Newsvendor under Ambiguity and Misspecification

Feng Liu[1], Zhi Chen[2], Ruodu Wang[3], and Shuming Wang[1]

[1]School of Economics and Management, University of Chinese Academy of Sciences
Emails: liufeng21@mails.ucas.ac.cn, wangshuming@ucas.edu.cn

[2]Department of Decisions, Operations and Technology, The Chinese University of Hong Kong
Email: zhi.chen@cuhk.edu.hk

[3]Department of Statistics and Actuarial Science, University of Waterloo
Email: wang@uwaterloo.ca

Problem definition: We consider a newsvendor problem with unknown demand distribution, where we distinguish *ambiguity* under which the newsvendor does not differentiate demand distributions of common characteristics (*e.g.*, mean and variance) and *misspecification* under which such characteristics might be misspecified (due to, *e.g.*, estimation error and/or distribution shift).

Methodology/results: The newsvendor hedges against ambiguity and misspecification by maximizing the worst-case expected profit regularized by a distribution's distance to an ambiguity set of distributions with some specified characteristics. Focusing on the popular mean-variance ambiguity set and optimal-transport cost for the misspecification, we show that the decision criterion of misspecification aversion possesses insightful interpretations as distributional transforms. We derive the closed-form optimal order quantity that generalizes the solution of the seminal Scarf model under only ambiguity aversion. We establish the finite-sample performance guarantee, which consists of two parts: in-sample optimal value and out-of-sample effect of misspecification that can be further decoupled into estimation error and distribution shift. We also extend the framework to multiple products, distributional characteristics specified via optimal transport, and misspecification measured by total variation distance, and derive analytical optimal solutions.

Managerial implications: The closed-form solution highlights the impact of misspecification aversion: the optimal order quantity under misspecification aversion can decrease as the price or variance increases, reversing the monotonicity of that under only ambiguity aversion. Hence, ambiguity and misspecification, as different layers of distributional uncertainty, can result in distinct operational consequences. The finite-sample performance guarantee theoretically justifies the necessity of incorporating misspecification aversion in a non-stationary environment, which is also well demonstrated in our experiments with real-world data.

Key words: newsvendor, model misspecification, moment condition, optimal transport, performance guarantee.

---

1. Introduction

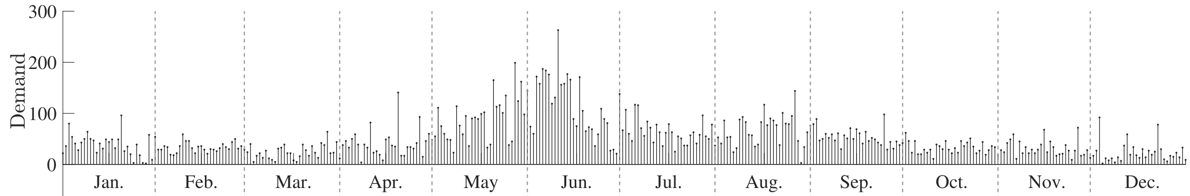
Newsvendor model, as a building block for dealing with uncertain demand in operations management, has found its successful applications in various domains, including inventory management (Chen et al. 2007), revenue management (Besbes et al. 2018), capacity planning (Simchi-Levi and Wei 2015), and healthcare (Olivares et al. 2008), to name a few. When the true demand distribution is fully known, the celebrated *critical fractile* determines an optimal order quantity that maximizes the expected profit. In practice, however, the newsvendor often faces incomplete knowledge about the demand distribution. Hence, it can be difficult (if not impossible) to precisely articulate the true demand distribution, leading to the issue of demand *ambiguity* to the newsvendors.

A natural way to mitigate demand ambiguity is to utilize only partial distributional characteristics available for decision-making. In this vein, mean and variance—arguably two of the most widely used and easy-to-estimate statistics that capture the key *location* and *dispersion* characteristics of the underlying distribution, respectively—have been employed. This can be traced back to the seminal work of Scarf (1958) that considers an *ambiguity-averse* newsvendor model maximizing the worst-case expected profit over an ambiguity set of probability distributions with the same mean and variance.[1] In essence, with such a mean-variance ambiguity set, the newsvendor specifies her belief about the (true) demand distribution via an approximation by using mean and variance characteristics. This is also well statistically justified,[2] especially in the context of the newsvendor problem, by noting that the single-dimensional demand distribution's quantile (*i.e.*, the critical-fractile solution) can be largely characterized by the mean and variance statistics, or even *perfectly* determined under many common distributions (*e.g.*, elliptical, uniform, and exponential, see Meyer 1987). This also underpins the implication that solutions of the ambiguity-averse and nominal ambiguity-neutral models characterized by the same mean and variance can share several key features in operational properties: for instance, the order quantity increases in price (see Sections 2 and 4.2 for more details). Apart from the newsvendor problem (Perakis and Roels 2008, Han et al. 2014), the mean-variance ambiguity set has been studied in stochastic optimization (Popescu 2007), and used in various applications, including recent literature on decision theory (Müller et al. 2022), mechanism design (Carrasco et al. 2018, Chen et al. 2022, 2024b), and risk management (Li et al. 2018, Nguyen et al. 2021).

However, in many practical situations, it is still challenging to accurately estimate the demand mean and variance. For instance, in retailing industries (*e.g.*, E-commerce, grocery, and supermarket), more and more commodities are of ever shorter life cycles (Calvo and Martínez-de Albéniz 2016), and many enterprises lack the resources for effective data collection and analysis, especially for new products (Feiler and Tong 2022). Insufficient historical data—even when the underlying demand process is stationary—can result in non-negligible *estimation error* that challenges the newsvendors' decision under demand ambiguity. On the other hand, the uncertainty is exacerbated

[1] A more general version of the Scarf model allows the mean and variance to be uncertain and to vary within some predetermined intervals, which turns out to be equivalent to the Scarf model characterized by the predetermined lower bound of mean and upper bound of variance (see Natarajan et al. 2011 or Remark 1).

[2] It is well known that under mild conditions, a probability distribution can be *uniquely* determined by all integer-order moments (Durrett 2019), among which the first two moments are the most commonly used ones.



	Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.
Mean	38.58	33.03 ✓	24.87 ✗	44.43 ✗	88.80 ✗	115.23 ✗	60.65 ✗	69.65 ✓	51.10 ✗	30.61 ✗	30.86 ✓	23.48 ✗
STD	21.33	9.32 ✗	14.38 ✗	29.87 ✗	43.14 ✗	58.99 ✗	21.95 ✗	29.53 ✗	14.00 ✗	9.92 ✗	15.89 ✗	20.95 ✓

Figure 1 Daily demand of a product of drinking water over one year with monthly mean and standard deviation (STD). The sign '✓' (resp., '✗') means that the monthly mean or STD falls in (resp., does not fall in) the 95% confidence interval estimated with the demand data in the preceding month. Notably, in only 2 out of 11 (resp., 1 out of 11) instances, the mean (resp., variance) falls within the confidence interval.

Notes. The confidence intervals on the mean and variance are constructed leveraging the t-statistic and χ^2-statistic, respectively, without knowing the true values of mean and variance.

if the underlying demand process becomes non-stationary, due to, for instance, the complicated (time-varying) determinants for the demand (Keskin and Zeevi 2017, An et al. 2025). This can lead to *distribution shift*—the future demand distribution is different from the past[3]—under which the mean and variance characteristics can constantly change over time, making them difficult to estimate from historical data. In Figure 1, we illustrate that even the confidence interval of either mean or variance may depreciate in a non-stationary demand environment. In sum, either estimation error or distribution shift can cause *misspecified* mean and variance parameters, and consequently, the optimal order quantity prescribed under ambiguity may perform inexpertly. This is known as *model misspecification* in statistics and economics (Hansen 2014). In our context, it refers to the possibility that the true demand distribution may not reside in the ambiguity set with the same distributional characteristics as specified (*e.g.*, mean and variance in the Scarf model or the neighborhood of some reference distribution as in Section 6.2); see Figure 2 for an illustration.

The above discussion motivates us to caution against the potential issue of model misspecification in the fundamental newsvendor problem. To this end, we introduce a misspecification-averse (and ambiguity-averse) newsvendor model, which is inspired by a recently developed framework of globalized distributionally robust optimization (Liu et al. 2023) that can mitigate the performance shortfall due to the possibility that the true distribution falls outside the ambiguity set. Our newsvendor model is also well supported by a recent axiomatic framework (Cerreia-Vioglio et al. 2025) that unifies behavioral decision criteria averse to either ambiguity or misspecification. In particular, we distinguish the ambiguity under which the newsvendor does *not* differentiate demand

[3] The phenomenon of distribution shift, in statistical learning, also refers to situations where the training and testing samples are governed by different distributions (see, *e.g.*, Quiñonero-Candela et al. 2008).

Figure 2 *Left*: The true demand distribution F resides in the ambiguity set $\mathcal{A}$, and the newsvendor faces only ambiguity. *Right*: The true demand distribution resides outside the ambiguity set, and hence, the newsvendor faces both ambiguity and misspecification.

distributions of common distributional characteristics (*e.g.*, mean and variance) and misspecification under which such characteristics might be misspecified (due to, *e.g.*, estimation error and/or distribution shift as discussed before). The newsvendor hedges against ambiguity and misspecification by maximizing the worst-case expected profit regularized by a distribution's distance to an ambiguity set of distributions with some specified characteristics. We investigate—from decision-criterion, operational, and statistical perspectives—how misspecification aversion may affect the newsvendor's decision and what is the rationale behind it.

1.1. Summary of Main Contributions

Newsvendor under misspecification. We investigate the fundamental newsvendor problem under a *structured* decision-under-uncertainty framework that distinguishes the ambiguity and misspecification. We focus on the mean-variance information for characterizing the ambiguity and optimal-transport cost for quantifying the misspecification, respectively, and investigate the rationale behind the misspecification aversion that differentiates from the ambiguity aversion of the seminal Scarf framework. We extend our modeling on the ambiguity of demand distribution via bounding its statistical distance to a reference distribution, misspecification aversion upon ambiguity-aversion is then equivalent to a stronger misspecification aversion to a nominal ambiguity-neutral model under the reference distribution (Section 6.2). We also extend to consider other statistical distances (specifically, total variation distance) to measure the misspecification (Section 6.3). These, together with misspecification aversion to the mean-variance ambiguity set, achieve a comprehensive inspection of the newsvendor under ambiguity and misspecification.

Interpretation via distributional transform. We investigate and interpret the decision criterion of misspecification aversion by leveraging a decision-analysis vehicle of *distributional transform* (Liu et al. 2021). We show that the decision criterion of the newsvendor under ambiguity and misspecification is essentially a worst-case *transformed* expected profit with each distribution *within* the ambiguity set being transformed—via a *transform function* determined by the index of misspecification aversion and operational information (*e.g.*, price p and order quantity q) of the newsvendor—to another distribution possibly *beyond* the ambiguity set (Theorem 1). In this way, the misspecification-aversion effect with valuable information on cost structure (*e.g.*, price) and ordering decision is encapsulated in the transform function along the distributional transform.

Analytical optimal solutions. We analytically derive the optimal order quantity of the newsvendor under ambiguity and misspecification (Theorem 2), which generalizes that of the seminal Scarf model under ambiguity alone. The analytical solution enables us to analyze the optimal order quantity's sensitivity to cost structure (*i.e.*, price and cost) and distributional characteristics (*i.e.*, mean and variance) to understand the impact and rationale of misspecification aversion. In particular, while it is *always* optimal to order *more* as price increases under ambiguity aversion, it turns out that ordering *less* instead can be optimal as the price increases under misspecification aversion, *ceteris paribus* (Proposition 2). Likewise, in the case of high-profit margin, while it is optimal to order *more* as variance increases under ambiguity aversion, ordering *less* can be optimal as the variance increases under misspecification aversion, *ceteris paribus* (Proposition 3). These observations reveal that ambiguity and misspecification, as different layers of distributional uncertainty, can result in *distinct* operational consequences, and therefore should be distinguished in the modeling. Furthermore, we extend to derive the analytical optimal order quantities of multiple products under a sum-of-variance constraint, which *unifies* the ambiguity-averse and misspecification-averse single-product model, ambiguity-averse multi-product model, and the Scarf model (Theorem 5).

Performance guarantee. We establish the finite-sample performance guarantee of the optimal solution that decouples the mixing effect of estimation error and distribution shift in the misspecification statistically (Theorem 3). In particular, the estimation error is related to the distance between the data-generating distribution and the estimated mean-variance ambiguity set, and it *diminishes* as the sample size approaches infinity; while the distribution shift, captured by the distance between the data-generating distribution and the out-of-sample distribution, is *independent* of the sample size. This theoretically justifies the *rationale* of incorporating misspecification aversion in a non-stationary environment, which is also well demonstrated in our experiments with real-world retailing data.

1.2. Related Works

Our work is related to the following streams of literature.

Newsvendor with moment condition. Since the pioneering work of Scarf (1958), various studies have employed the moment condition to specify the ambiguity-averse newsvendor model with limited demand information. For instance, building on the results of Scarf (1958), Gallego and Moon (1993) extend to consider a multi-product newsvendor problem based on *marginal* mean-and-variance information under a budget constraint. Natarajan et al. (2011) generalize the Scarf model by allowing the mean and variance to be uncertain and to vary within some predetermined intervals. This formulation turns out to be equivalent to the Scarf model characterized by the lower bound of mean and the upper bound of variance. Zhu et al. (2013) focus on minimizing the worst-case regret under known mean and variance of the random demand. In addition to ambiguity aversion, Han et al. (2014) incorporate risk aversion into the newsvendor problem, where the risk is captured by the standard deviation of the newsvendor's profit. Given the mean and variance of the demand, they develop a closed-form solution for the risk-averse (and ambiguity-averse) newsvendor. Apart from mean and variance, other moment information has also been considered in ambiguity-averse newsvendor problems. We refer to Perakis and Roels (2008) for structural information such as median, unimodality, and symmetry, to Ardestani-Jaafari and Delage (2016) for first-order partial moments, to Natarajan et al. (2018) for asymmetry based on second-order partitioned statistics, to Das et al. (2021) for the t-th ($t \geq 1$) moment that can capture heavy-tailed demand distributions, and to Padmanabhan et al. (2021) for partial correlations among products. Specifically, Das et al. (2021) interpret the Scarf model through a heavy-tailed demand perspective, and reveal that the robust newsvendor problem can be represented under an implied heavy-tailed distribution outside the moment ambiguity set. Govindarajan et al. (2021) shift the focus to the inventory pooling problem, where the ambiguity set is specified by mean and covariance (see also Hanasusanto et al. 2015), and characterize the closed-form solution of the two-location model.

Newsvendor with statistical distance. Another stream of the ambiguity-averse newsvendor model is based on ambiguity sets specified through the closeness to a reference distribution in terms of a certain statistical distance. For instance, Rahimian et al. (2019) delve into the total variation distance and obtain an insightful closed-form solution. Based on the Wasserstein distance, Chen and Xie (2021) adopt the minimax regret decision criterion in the presence of both demand and yield randomness, and they show that the optimal order quantity can be determined via an efficient golden section search. Zhang et al. (2023) and Fu et al. (2024) further leverage side information from explanatory features, and derive an optimal analytical ordering policy based on the Wasserstein distance and a closed-form solution based on the JW discrepancy measure, respectively. In this work, we also consider the possibility that the misspecification may arise from such ambiguity-averse newsvendor models with statistical distance (Theorem 6).

Model misspecification. Statisticians and econometricians have long grappled with the challenge of addressing uncertainty in decision-making, which is categorized as risk, ambiguity, and misspecification (see, *e.g.*, Hansen 2014, Hansen and Sargent 2022). Several noteworthy contributions have been made, which, from many angles, shed light on the interplay between the different layers of uncertainty. From the empirical perspective, Aydogan et al. (2023) provide experimental evidence for the role of model misspecification in decision-making under uncertainty. Their work establishes a compelling case for recognizing a distinct behavioral impact among risk, ambiguity, and misspecification, illuminating the nuanced facets of decision theory. From the theoretical perspective, Cerreia-Vioglio et al. (2025) propose an innovative axiomatic framework that unifies the behavioral decision criteria on the aversion to ambiguity and/or misspecification. From the perspective of optimization under uncertainty, the works of Ben-Tal et al. (2006) and Ben-Tal et al. (2017) propose paradigms named globalized/comprehensive robust counterparts to mitigate the violation of uncertainty-involved constraints when the true parameter does not reside in the pre-specified uncertainty set. Recently, Liu et al. (2023) extend the notion of globalized robustness and propose the globalized distributionally robust counterpart to mitigate the violation when the true distribution does not reside in the pre-specified ambiguity set. In addition, Long et al. (2023) introduce a target-oriented model of robust satisficing that maximizes the robustness to uncertainty of achieving a satisfactory target. Also, the distributionally robust models of Delage and Ye (2010) and Wiesemann et al. (2014) can mitigate the possible variations in the moment information and in the probabilities of specific events, respectively. From the operational perspective, De Meyer et al. (2002) delineate four types of uncertainty in project management, namely variation, foreseen uncertainty, unforeseen uncertainty, and chaos. The latter two align closely with the notions of ambiguity and misspecification. There are also nascent works investigating the issue of model misspecification in revenue management, such as a misspecified demand function (Nambiar et al. 2019) and a misspecified choice model (Chen et al. 2024a).

1.3. Research Gaps

Although our work is closely related to the studies of Cerreia-Vioglio et al. (2025) and Liu et al. (2023), the gap can be well established. It is noted that Cerreia-Vioglio et al. (2025) focuses primarily on developing the decision-theoretical axioms for the model misspecification, and Liu et al. (2023) develops the framework of globalized distributionally robust optimization and focuses on deriving the associated tractable model counterparts. However, our study is devoted to the analysis of the newsvendor problem under both ambiguity and misspecification. In particular, our developments (for the newsvendor problem)—on the decision-criterion interpretation from the perspective of distributional transform, closed-form and analytical solutions, operational insights, and the performance guarantee that decouples the effects of estimation error and distribution shift—have not been attempted in the above studies. We also point out that although the ambiguity set proposed in Delage and Ye (2010) also captures the variation in mean and variance, it is, however, computationally inconvenient for the newsvendor problem. Moreover, our approach can be readily adapted—by replacing the Scarf ambiguity set with theirs—to further mitigate the misspecification to the bounds on the moment information in their ambiguity set. Although both our work and Das et al. (2021) admit an equivalent stochastic representation under an implied distribution that lies outside the prescribed ambiguity set, our model captures misspecification aversion, whereas Das et al. (2021) reinterpret the Scarf model via the lens of heavy-tailed demand.

1.4. Notation

We denote by $\mathcal{P}$ (resp., $\mathcal{M}_+$) the set of probability measures (resp., non-negative measures) supported on $\mathbb{R}_+$, and $\mathcal{P}_0$ the set of probability measures supported on $\mathbb{R}$. We use $\tilde{v} \sim F$ to signify a random variable $\tilde{v}$ that follows the probability distribution with a cumulative probability distribution (CDF) F, under which the expectation is $\mathbb{E}_F[\cdot]$. The Dirac distribution at $v \in \mathbb{R}$ is denoted by δ_v. We adopt the convention that $1/0 = \infty$, and refer to "decreasing/increasing" in the weak sense.

2. Model

The newsvendor decides the order quantity before demand realization and tries to maximize her expected profit. Given the unit price p, unit cost c ($c < p$), and an order quantity q, the profit under the materialized demand v amounts to $\pi(q,v) = p \cdot \min\{q,v\} - cq = p \cdot \min\{q,v\} - (1-\kappa)pq$, where we denote the profit margin by $\kappa = \frac{p-c}{p}$ (giving $c = (1-\kappa)p$), which will be an important parameter for our analysis. Facing stochastic demand, the newsvendor must navigate optimizing her decision-making process to balance the trade-off between lost sales (incurred when $v > q$) and excess inventory (incurred when $v \leq q$). When the precise demand distribution G is fully known, the newsvendor maximizes her expected profit by solving the problem

$$\max_{q \geq 0} \; \mathbb{E}_G[\pi(q, \tilde{v})]. \tag{Nominal}$$

The optimal order quantity $q^\star_G$ is characterized by the classic critical fractile $q^\star_G = G^{-1}(\kappa)$ (Arrow et al. 1951). If G is an elliptical distribution with location parameter μ, scale parameter σ, and some density generator $\xi(\cdot)$,[4] then $q^\star_G = \mu + \sigma \cdot \Xi^{-1}(\kappa)$, where $\Xi(u) = \int_{-\infty}^{u} \xi(v^2)\mathrm{d}v$. Such a mean-variance formula holds for many other classes of distributions, including the uniform distributions and the exponential distributions. It is worth mentioning that within each class above, the optimal order quantity is also uniquely determined by only the mean and variance of the demand distribution.

In practice, full information on the demand distribution is often not accessible. To tackle the challenge of partial distributional information, significant advancements have been made in prescribing an ambiguity-averse solution that remains robust against all distributions specified by some common distributional characteristics. The seminal work of Scarf (1958) specifies only the mean and variance of the demand distribution and solves

$$\max_{q \geq 0} \min_{G \in \mathcal{A}} \mathbb{E}_G[\pi(q, \tilde{v})] \quad \text{with} \quad \mathcal{A} = \{G \in \mathcal{P} \mid \mathbb{E}_G[\tilde{v}] = \mu, \, \mathbb{E}_G[\tilde{v}^2] = \mu^2 + \sigma^2\}. \tag{Ambiguity}$$

[4] The density function of the elliptical distribution is $f(x) = \frac{1}{\sigma} \cdot \xi\big(\frac{(x-\mu)^2}{\sigma^2}\big)$ with $\int_{-\infty}^{\infty} \xi(z^2)\,\mathrm{d}z = 1$.

The Ambiguity model admits an analytical solution—for the reason that will become clear subsequently, we emphasize it with a subscript '∞'—as follows:

$$q^\star_\infty = \begin{cases} \mu + \sigma f(1-\kappa) & \kappa \geq \dfrac{\sigma^2}{\mu^2+\sigma^2} \\ 0 & \kappa < \dfrac{\sigma^2}{\mu^2+\sigma^2} \end{cases} \quad \text{with} \quad f(x) = \frac{1-2x}{2\sqrt{x(1-x)}} \quad \forall 0 < x < 1. \tag{1}$$

As illustrated in Scarf (1958), the optimal order quantity $q^\star_\infty$ of Ambiguity is comparatively close to the optimal order quantity of Nominal under a normal approximation of the Poisson distribution for a moderate range of profit margins. In effect, in many commonly used distributions (*e.g.*, elliptical, uniform, and exponential) of a nominal G as mentioned above, the optimal order quantities $q^\star_\infty$ and $q^\star_G$ share important sensitivity features (to be discussed in Section 4.2). In addition, the Ambiguity model can also cover the situation where the mean and variance are themselves uncertain and reside in some estimated intervals, as remarked below.

Remark 1 (Generality of Ambiguity). Consider an ambiguity set parameterized by bounds on the uncertain mean and variance:

$$\mathcal{V} = \left\{ G \in \mathcal{P} \mid \mathbb{E}_G[\tilde{v}] = \mu,\, \mathbb{E}_G[\tilde{v}^2] = \mu^2 + \sigma^2 \ \text{ for some } \ \mu \in [\underline{\mu}, \bar{\mu}] \ \text{ and } \ \sigma^2 \in [\underline{\sigma}^2, \bar{\sigma}^2] \right\}.$$

By corollary 5.1 of Natarajan et al. (2011), it holds that

$$\max_{q\geq 0} \min_{G\in\mathcal{V}} \mathbb{E}_G[\pi(q,\tilde{v})] = \max_{q\geq 0} \min_{G\in\mathcal{A}} \mathbb{E}_G[\pi(q,\tilde{v})] \ \text{ with } \ \mathcal{A} = \{G \in \mathcal{P} \mid \mathbb{E}_G[\tilde{v}] = \underline{\mu},\, \mathbb{E}_G[\tilde{v}^2] = \underline{\mu}^2 + \bar{\sigma}^2\}.$$

The right-hand side equivalence is indifferent to Ambiguity with mean $\underline{\mu}$ and variance $\bar{\sigma}^2$. Similarly, if the moment ambiguity set developed in Delage and Ye (2010) is employed, the bounds on the mean and variance could also be misspecified.

An underlying assumption of Ambiguity is that the true distribution F_{true} falls in the specified ambiguity set $\mathcal{A}$. However, the ambiguity set $\mathcal{A}$ is constructed using the mean and variance estimates for those of the true F_{true}, and is, therefore, may be misspecified, that is, $F_{\text{true}} \notin \mathcal{A}$ (see also the empirical evidence in Figure 1). This leads to *misspecification*—an issue we address next.

To capture misspecification formally and in a computationally appealing fashion,[5] for a demand distribution $F \in \mathcal{P}$, we measure its closeness to the ambiguity set $\mathcal{A}$ by

$$d(F,\mathcal{A}) = \min_{G\in\mathcal{A}} d(F,G)$$

[5] When considering misspecification aversion, Cerreia-Vioglio et al. (2025) focus on ϕ-divergence, which is generally computationally difficult in the newsvendor problem, except for the total variation distance as discussed in Section 6.3. Another alternative is the Gelbrich distance (Gelbrich 1990), which, however, does not lead to an analytical solution.

with $d(F,G)$ being the optimal-transport cost (Villani 2009) between two distributions F and G with quadratic cost function $|\cdot|^2$ defined as

$$d(F,G) = \min_{\Gamma \in \mathcal{W}(F,G)} \int_{\mathbb{R}_+ \times \mathbb{R}_+} |u-v|^2 \mathrm{d}\Gamma(u,v), \tag{2}$$

where $\mathcal{W}(F,G)$ is the set of joint probability distributions on $\mathbb{R}_+ \times \mathbb{R}_+$ with marginals F and G. The quantity $\sqrt{d(F,G)}$ is also known as the type-2 Wasserstein distance between F and G. Note that $d(F,\mathcal{A}) > 0$ if and only if $F \notin \mathcal{A}$. In the main content, we focus on the optimal-transport cost in characterizing misspecification, for the sake of tractability and statistical convenience (see more details in Section 4 and Section 5). Following the spirit of Cerreia-Vioglio et al. (2025), given the ambiguity set $\mathcal{A}$, we incorporate misspecification into the newsvendor's decision criterion so that a misspecification-averse (and ambiguity-averse) newsvendor solves

$$\Pi^\star_\alpha = \max_{q\ge 0} \min_{F\in\mathcal{P}} \left\{ \mathbb{E}_F[\pi(q,\tilde{u})] + \alpha \cdot d(F,\mathcal{A}) \right\} \tag{Misspecification}$$

for some $\alpha \ge 0$ that represents the index of misspecification aversion: the lower the index, the stronger the aversion to misspecification. Intuitively speaking, a larger value of α puts a larger penalty on deviation from the ambiguity set $\mathcal{A}$ (as measured by $d(F,\mathcal{A})$) and corresponds to higher confidence in $\mathcal{A}$ (or equivalently, the Ambiguity model). On the one end, when $\alpha \to \infty$, misspecification aversion is absent (see section 4.1 in Cerreia-Vioglio et al. 2025), and Misspecification reduces to Ambiguity as the newsvendor is absolutely confident with $\mathcal{A}$. On the other end, when $\alpha \to 0$, misspecification aversion is strongest, and Misspecification reduces to the robust model

$$\max_{q\ge 0} \min_{F\in\mathcal{P}} \mathbb{E}_F[\pi(q,\tilde{u})] = \max_{q\ge 0} \min_{u\in\mathbb{R}_+} \pi(q,u),$$

wherein the newsvendor is so unconfident that she disregards the distributional characteristics specified in $\mathcal{A}$. An analogous observation to Remark 1 for Ambiguity can be established for Misspecification: misspecification over the uncertain mean and variance is still equivalent to misspecification over the lowest mean and highest variance; see Section EC.3.

It is worth noting that the Misspecification problem can also be regarded as the Lagrangian relaxation of an alternate formulation of the newsvendor model under misspecification as follows:

$$\max_{q\ge 0} \min_{d(F,\mathcal{A})\le\varepsilon} \mathbb{E}_F[\pi(q,\tilde{u})] \tag{3}$$

for some $\varepsilon \ge 0$. When $\varepsilon = 0$, problem (3) reduces to Ambiguity; and when $\varepsilon > 0$, the optimal solution to problem (3) can be constructed from that of Misspecification (Lemma EC.2, Section EC.1). In other words, the decision of Misspecification essentially hedges against some worst-case distribution that can stay *outside* the ambiguity set $\mathcal{A}$ characterized by mean and variance. This makes Misspecification distinct from Ambiguity (with the underlying worst-case distribution within $\mathcal{A}$) in many aspects, which we further explore in the forthcoming sections. In Section 6, we extend Misspecification to involve multiple products, distance-based ambiguity set defined via optimal transport, and misspecification measured by total variation distance.

3. Decision Criterion via Distributional Transform

To understand the decision criterion of misspecification aversion in our newsvendor context, we investigate the objective function of Misspecification from a perspective of *distributional transform*. We show that the objective function essentially transforms distributions in $\mathcal{A}$—via a *transform function* determined by the index α of misspecification aversion and the newsvendor's profit function—to new ones that possibly violate the mean and variance constraints specified by $\mathcal{A}$.

Exploring the definition of $d(F,\mathcal{A})$ and interchanging the minimization over F and G, we can represent equivalently the Misspecification problem as follows:

$$\max_{q\geq 0}\min_{G\in\mathcal{A}}\min_{F\in\mathcal{P}}\Big\{\mathbb{E}_F[\pi(q,\tilde{u})]+\alpha\cdot d(F,G)\Big\}.$$

Here, the inner minimization features the potential misspecification of a fixed distribution $G\in\mathcal{A}$, which is then robustified over the specified ambiguity set $\mathcal{A}$ via the outer minimization. Recall that a distributional transform $T_\varphi[\cdot]:\mathcal{P}\mapsto\mathcal{P}_0$ maps distributions in $\mathcal{P}$ to $\mathcal{P}_0$ via a *transform function* φ (Liu et al. 2021). We can then represent the innermost minimization above through a distributional transform, leading to the following result that plays a key role in characterizing the decision criterion of Misspecification. Different from Liu et al. (2021), the transform function we identify further encodes operational information such as price parameter p and order quantity q.

Theorem 1 (Distributional Transform). *Given $\alpha\geq 0$ and $q\geq 0$, it holds that*

$$\min_{F\in\mathcal{P}}\Big\{\mathbb{E}_F[\pi(q,\tilde{u})]+\alpha\cdot d(F,\mathcal{A})\Big\}=\min_{G\in\mathcal{A}}\int_{\mathbb{R}_+}\pi(q,v)\,\mathrm{d}T_{\varphi_\alpha}[G](v),$$

where $T_{\varphi_\alpha}[G](v)=G\circ\varphi_\alpha^{-1}(v)\ \forall v\in\mathbb{R}_+$, and $T_{\varphi_\alpha}[\cdot]$ is a distributional transform of $G\in\mathcal{A}$ with an increasing and continuous transform function $\varphi_\alpha:\mathbb{R}_+\mapsto\mathbb{R}_+$ defined as follows.

(i) If $\alpha<\frac{p}{4q}$, then $\varphi_\alpha(v)=\frac{\alpha}{p}\cdot v^2$.

(ii) If $\alpha\geq\frac{p}{4q}$, then $\varphi_\alpha(v)=\frac{\alpha}{p}\cdot v^2$ when $v<\frac{p}{2\alpha}$ and $\varphi_\alpha(v)=v-\frac{p}{4\alpha}$ otherwise.

By Theorem 1, the decision criterion of Misspecification serves as an expectation under a transformed worst-case distribution (Figure 3), which establishes the equivalence between Misspecification and the following problem

$$\max_{q\geq 0}\min_{G\in\mathcal{A}}\mathbb{E}_{T_{\varphi_\alpha}[G]}[\pi(q,\tilde{v})].\qquad\text{(Transform)}$$

For the remainder of this section, we may use Transform and Misspecification interchangeably.

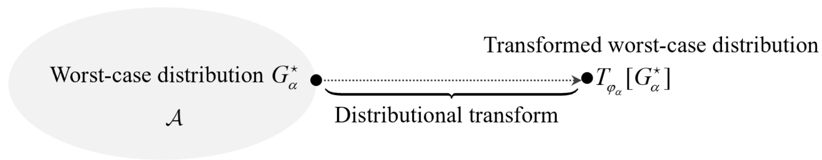


Figure 3 The decision criterion of MISSPECIFICATION transforms the worst-case distribution $G^{\star}_{\alpha}$ in the ambiguity set $\mathcal{A}$ into the transformed worst-case distribution $T_{\varphi_\alpha}[G^{\star}_{\alpha}]$ that can be outside of $\mathcal{A}$.

Recall that AMBIGUITY evaluates the performance of an order quantity q via the decision criterion

$$\min_{G\in\mathcal{A}} \mathbb{E}_G[\pi(q,\tilde{v})].$$

The above equivalence reveals that the key *difference* between AMBIGUITY and MISSPECIFICATION lies in the transform function φ_α applied to the probability distributions in the mean-variance ambiguity set $\mathcal{A}$. In particular, for any *physical*[6] probability distribution $G\in\mathcal{A}$, φ_α transforms it to a *subjective* probability distribution $T_{\varphi_\alpha}[G]$ that can be *outside* $\mathcal{A}$, so that the resulting transformed expectation $\mathbb{E}_{T_{\varphi_\alpha}[G]}[\cdot]$ reflects the newsvendor's aversion to misspecification. On the one hand, for a small value of α such that $\alpha<\frac{p}{4q}$, the transform function φ_α compresses (resp., amplifies) low (resp., high) demand realizations of G,[7] and a smaller α results in more demand realizations being compressed, see the cases of α_1,α_2 on the left panel of Figure 4. On the other hand, for a large value of α such that $\alpha\geq\frac{p}{4q}$, φ_α compresses all demand realizations of G;[8] see the case of α_3 on the left panel of Figure 4. In this case, misspecification aversion compresses all probability distributions in $\mathcal{A}$, and a smaller α also leads to a stronger compression. Importantly, due to the transform function φ_α, the mean or variance of the transformed distribution $T_{\varphi_\alpha}[G]$ can be different from that of the original distribution G in the ambiguity set $\mathcal{A}$, namely $T_{\varphi_\alpha}[G]\notin\mathcal{A}$; see the right panel of Figure 4 for a visualization. As $\alpha\to\infty$, the transform function becomes $\varphi_\alpha(v)=v$, implying $T_{\varphi_\alpha}[G]=G$ and that MISSPECIFICATION reduces to AMBIGUITY. That is, the index α of misspecification aversion, or equivalently, the newsvendor's aversion against misspecification of $\mathcal{A}$, is fully encoded in the transform function φ_α of TRANSFORM. The transformed distribution $T_{\varphi_\alpha}[G]$ could be far away from the ambiguity set $\mathcal{A}$, especially when α is small such that the mean of $T_{\varphi_\alpha}[G]$ is significantly lower than that of distributions in $\mathcal{A}$ (see the case of α_1 on the right panel of Figure 4).

Interestingly, we emphasize that given α and an order quantity q, the distributional transform T_{φ_α} is determined by the price p (see the transform function φ_α derived in Theorem 1), leading

[6] All distributions in $\mathcal{A}$ sharing the same *physically* observable mean-variance information are treated indifferently.

[7] When $\alpha<\frac{p}{4q}$, $T_{\varphi_\alpha}[G](v)=G(\sqrt{\frac{p}{\alpha}v})$ for every $G\in\mathcal{A}$. Since $G(v)$ is increasing in v, it can be seen that $T_{\varphi_\alpha}[G](v)>G(v)$ when $v<\frac{p}{\alpha}$ and $T_{\varphi_\alpha}[G](v)\leq G(v)$ when $v\geq\frac{p}{\alpha}$.

[8] When $\alpha\geq\frac{p}{4q}$, $T_{\varphi_\alpha}[G](v)=G(\sqrt{\frac{p}{\alpha}v})\geq G(v)$ if $v<\frac{p}{2\alpha}$ and $T_{\varphi_\alpha}[G](v)=G(v+\frac{p}{4\alpha})\geq G(v)$ otherwise. Hence, for every $G\in\mathcal{A}$, we always have $T_{\varphi_\alpha}[G](v)\geq G(v)$ for all $v\geq 0$.

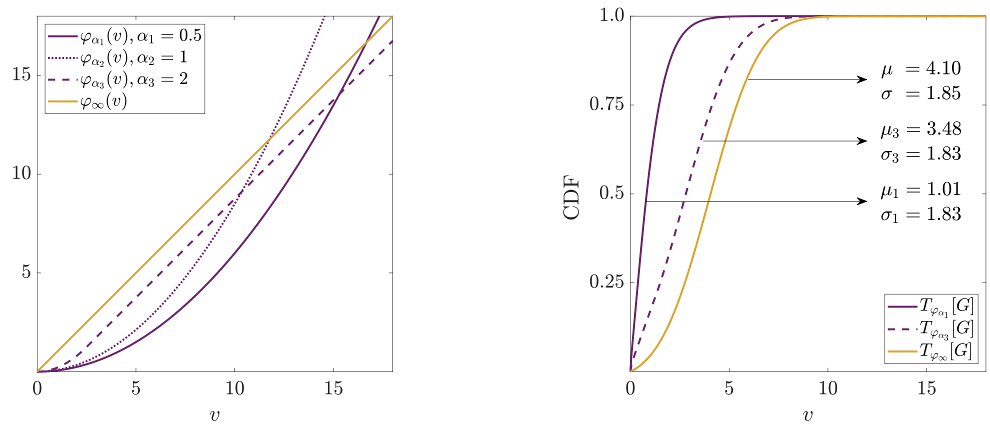


Figure 4 *Left*: Transform function $\varphi_\alpha(v)$. *Right*: CDF of a truncated normal distribution G and the transformed distribution $T_{\varphi_\alpha}[G]$. On both panels, $\alpha_1 < \alpha_2 < \frac{p}{4q} < \alpha_3$. For the right panel, $G \in \mathcal{A}$ is a normal distribution truncated to $\mathbb{R}_+$ with mean μ and standard deviation σ. Here, μ_1 and σ_1 (resp., μ_3 and σ_3) are the mean and standard deviation of the transformed distribution $T_{\varphi_{\alpha_1}}[G]$ (resp., $T_{\varphi_{\alpha_3}}[G]$).

to the transformed worst-case distribution $T_{\varphi_\alpha}[G^\star_\alpha]$ with $G^\star_\alpha \in \arg\min_{G\in\mathcal{A}} \mathbb{E}_{T_{\varphi_\alpha}[G]}[\pi(q,\tilde{v})]$ being dependent on the price (see Proposition EC.1 in Section EC.4). In contrast, we recall that the worst-case distribution implied by Ambiguity, *i.e.*, $G^\star_\infty \in \arg\min_{G\in\mathcal{A}} \mathbb{E}_G[\pi(q,\tilde{v})]$, is however independent of the cost structure.[9] In other words, the *price-independent* worst-case distribution *inside* the ambiguity set $\mathcal{A}$ in the decision criterion of Ambiguity now becomes a *price-dependent* transformed worst-case distribution *outside* of $\mathcal{A}$ in the decision criterion of Misspecification. We emphasize that such price-dependency effect indeed leads to operational consequences of Misspecification being distinct from that of Ambiguity, which will be further explored in Section 4.

4. Optimal Solution and Sensitivity Analysis

In this section, we first derive the optimal order quantity of Misspecification in closed form. We then investigate the impact of misspecification aversion via the optimal order quantity's sensitivity to the cost-structure information and distributional information, revealing important operational implications of Misspecification distinct from Ambiguity.

4.1. Analytical Solution

To proceed, we recall that Misspecification can be reformulated as

$$\max_{q\geq 0} \min_{F\in\mathcal{P}} \{\mathbb{E}_F[\pi(q,\tilde{u})] + \alpha\cdot d(F,\mathcal{A})\} = \max_{q\geq 0} \min_{G\in\mathcal{A}} \mathbb{E}_{T_{\varphi_\alpha}[G]}[\pi(q,\tilde{v})] = \max_{q\geq 0} \min_{G\in\mathcal{A}} \mathbb{E}_G[\pi(q,\varphi_\alpha(\tilde{v}))],$$

where the second equality follows from the definition of distributional transform T_{φ_α} as identified in Theorem 1. Hence, one can tackle Misspecification by adapting the primal-dual machinery

[9] Given an order quantity q, it can be shown that $G^\star_\infty = (\frac{\sigma^2}{\mu^2+\sigma^2})\cdot\delta_0 + (\frac{\mu^2}{\mu^2+\sigma^2})\cdot\delta_{\frac{\mu^2+\sigma^2}{\mu}}$ when $q < \frac{\mu^2+\sigma^2}{2\mu}$, and $G^\star_\infty = (\frac{1}{2}+\frac{q-\mu}{2w})\cdot\delta_{q-w} + (\frac{1}{2}-\frac{q-\mu}{2w})\cdot\delta_{q+w}$ otherwise, where $w = \sqrt{(q-\mu)^2+\sigma^2}$, which is independent of p.

for solving a maximin problem similar to Ambiguity but with a new "profit" function $\Psi(\alpha, q, v) = \pi(q, \varphi_\alpha(v))$. Specifically, given $q \geq 0$, the primal and dual reformulation of the inner worst-case expectation can be written as

$$\begin{array}{lll} \min\limits_{G \in \mathcal{M}_+} & \displaystyle\int_{\mathbb{R}_+} \Psi(\alpha, q, v)\, \mathrm{d}G(v) & \\ \text{s.t.} & \displaystyle\int_{\mathbb{R}_+} v\, \mathrm{d}G(v) = \mu & \cdots\ s_\alpha \\ & \displaystyle\int_{\mathbb{R}_+} v^2\, \mathrm{d}G(v) = \mu^2 + \sigma^2 & \cdots\ r_\alpha \\ & \displaystyle\int_{\mathbb{R}_+} \mathrm{d}G(v) = 1 & \cdots\ t_\alpha, \end{array} \iff \begin{array}{ll} \max\limits_{s_\alpha, r_\alpha, t_\alpha} & \mu s_\alpha - (\mu^2 + \sigma^2) r_\alpha - t_\alpha \\ \text{s.t.} & v s_\alpha - v^2 r_\alpha - t_\alpha \leq \Psi(\alpha, q, v) \quad \forall v \geq 0 \\ & s_\alpha \in \mathbb{R},\ r_\alpha \in \mathbb{R},\ t_\alpha \in \mathbb{R}. \end{array} \tag{4}$$

where s_α, r_α, and t_α are dual variables associated with the mean, variance, and support of the ambiguity set $\mathcal{A}$, respectively. The key to the primal-dual machinery is to construct a pair of primal and dual solutions that share identical objective values. In particular, the primal solution is a worst-case distribution constructed by identifying tangent points between the function $s_\alpha v - r_\alpha v^2 - t_\alpha$ and the function $\Psi(\alpha, q, v)$ in the dual reformulation. For Misspecification, however, the new "profit" function Ψ—neither convex nor concave in v—is less structured than the concave piecewise affine function π of Ambiguity, making the primal-dual procedure more involved to analyze. Fortunately, leveraging the closed form of $\Psi(\alpha, q, v) = \pi(q, \varphi_\alpha(v))$ given by Theorem 1, we can derive an analytical reformulation of the objective function in Misspecification.

Proposition 1 (Worst-Case Transformed Expectation). *Given $\alpha \geq 0$ and $q \geq 0$,*

$$\min_{G \in \mathcal{A}} \mathbb{E}_{T_{\varphi_\alpha}[G]}[\pi(q, \tilde{v})] = \begin{cases} \dfrac{p}{2}\left(q + \mu - \dfrac{p}{4\alpha} - \sqrt{\left(q - \mu + \dfrac{p}{4\alpha}\right)^2 + \sigma^2}\right) - cq & \text{if } q \in \mathcal{Q} \\ \dfrac{\alpha}{2}\left(\dfrac{pq}{\alpha} + \mu^2 + \sigma^2 - \sqrt{\left(\dfrac{pq}{\alpha} + \mu^2 + \sigma^2\right)^2 - 4\mu^2 \dfrac{pq}{\alpha}}\right) - cq & \text{otherwise,} \end{cases} \tag{5}$$

where $\mathcal{Q} = \{q \in \mathbb{R}_+ \mid q \geq \frac{p}{4\alpha},\ (2\mu - \frac{p}{\alpha})q \geq \mu^2 + \sigma^2 - \frac{p\mu}{2\alpha}\}$.

The analytical form of the worst-case transformed expectation is non-trivial and generalizes the worst-case expected cost of Ambiguity as $\alpha \to \infty$. Indeed, as $\alpha \to \infty$, $\mathcal{Q}$ becomes $\{q \in \mathbb{R}_+ \mid q \geq \frac{\mu^2+\sigma^2}{2\mu}\}$ and (5) recovers the worst-case expected cost of Ambiguity such that $\min_{G \in \mathcal{A}} \mathbb{E}_G[\pi(q, \tilde{v})] = \frac{\mu^2 pq}{\mu^2+\sigma^2} - cq$ if $q \geq \frac{\mu^2+\sigma^2}{2\mu}$ and $\min_{G \in \mathcal{A}} \mathbb{E}_G[\pi(q, \tilde{v})] = \frac{p}{2}(q + \mu - \sqrt{(q-\mu)^2 + \sigma^2}) - cq$ otherwise. Equipped with the analytical form, we can then derive the optimal solution of Misspecification as follows.

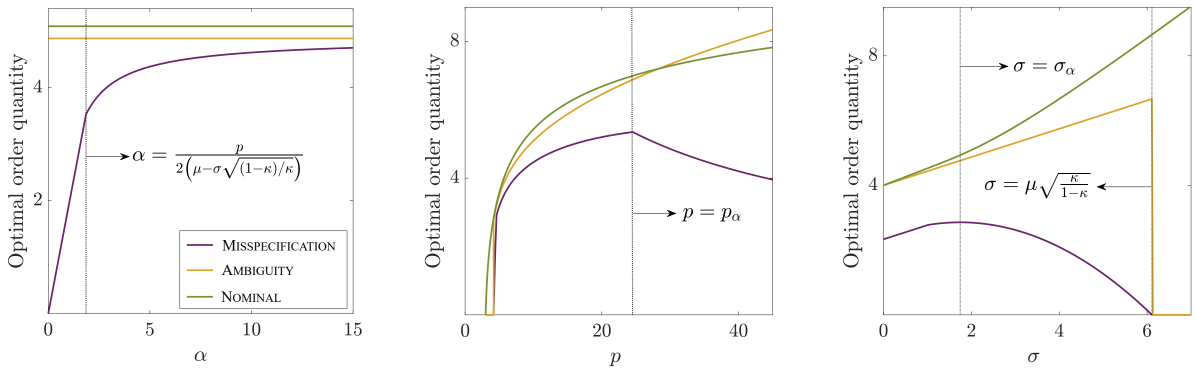


Figure 5 Optimal order quantity as a function of α (left), p (middle), and σ (right), respectively.

Notes. The optimal order quantity of Nominal is obtained under a normal distribution truncated to $\mathbb{R}_+$ with mean μ and standard deviation σ. For all three panels, we set $c=3$ and $\mu=4$. On the left panel, $p=10$ and $\sigma=2$. On the middle panel, $\sigma=2.5$ and $\alpha=4$, and we identify $p_\alpha=22.5$. On the right panel, $p=10$ and $\alpha=1.5$, for which $\kappa=0.7$ and the non-degenerate region is $\sigma\in\left[0,\mu\sqrt{\frac{\kappa}{1-\kappa}}\right]=\left[0,4\sqrt{\frac{7}{3}}\right]$, and we identify $\sigma_\alpha=\frac{8}{\sqrt{21}}$.

Theorem 2 (Optimal Solution). *Given $\alpha\geq 0$, the optimal order quantity $q^\star_\alpha$ of* Misspecification *is*

$$q^\star_\alpha=\begin{cases}\mu+\sigma f(1-\kappa)-\dfrac{p}{4\alpha} & \kappa\geq\dfrac{\sigma^2}{\mu^2+\sigma^2},\ \alpha\geq\dfrac{p}{2(\mu-\sigma\sqrt{(1-\kappa)/\kappa})}\\ \left(\mu^2-\sigma^2+2\mu\sigma f(1-\kappa)\right)\cdot\dfrac{\alpha}{p} & \kappa\geq\dfrac{\sigma^2}{\mu^2+\sigma^2},\ \alpha<\dfrac{p}{2(\mu-\sigma\sqrt{(1-\kappa)/\kappa})}\\ 0 & \kappa<\dfrac{\sigma^2}{\mu^2+\sigma^2},\end{cases}\tag{6}$$

where $f(\cdot)$ is defined in (1). The optimal order quantity $q^\star_\alpha$ is increasing in α.

Focusing on the non-degenerate case that $\kappa\geq\frac{\sigma^2}{\mu^2+\sigma^2}$, for $0\leq\alpha_1<\frac{p}{2(\mu-\sigma\sqrt{(1-\kappa)/\kappa})}\leq\alpha_2$, we have

$$q^\star_{\alpha_1}=\left(\mu^2-\sigma^2+2\mu\sigma f(1-\kappa)\right)\cdot\frac{\alpha_1}{p}\leq q^\star_{\alpha_2}=\mu+\sigma f(1-\kappa)-\frac{p}{4\alpha_2}\leq\mu+\sigma f(1-\kappa)=q^\star_\infty;$$

see left panel of Figure 5. This implies that the optimal order quantity $q^\star_\alpha$ of Misspecification is no larger than that of Ambiguity (*i.e.*, $q^\star_\infty$)—an intuitive result due to the additional aversion to misspecification—and $q^\star_\alpha\to q^\star_\infty$ as $\alpha\to\infty$. It is also notable that the optimal order quantity $q^\star_\alpha$ is affected by misspecification aversion and ambiguity aversion *separately*. When $\alpha<\frac{p}{2(\mu-\sigma\sqrt{(1-\kappa)/\kappa})}$, $q^\star_\alpha$ is a product of $\mu^2-\sigma^2+2\mu\sigma f(1-\kappa)$ and α/p that are purely determined by the mean and variance information specified in $\mathcal{A}$ and purely determined by misspecification, respectively. When $\alpha\geq\frac{p}{2(\mu-\sigma\sqrt{(1-\kappa)/\kappa})}$, $q^\star_\alpha$ is obtained by $\mu+\sigma f(1-\kappa)$ that is exactly the optimal order quantity $q^\star_\infty$ of Ambiguity minus $\frac{p}{4\alpha}$ that is purely determined by misspecification.

4.2. Sensitivity and its Implications

We next look at the optimal order quantity $q_\alpha^\star$'s sensitivity to the cost-structure information (*i.e.*, c and p) and distributional information (*i.e.*, μ and σ^2). Since it is straightforward that $q_\alpha^\star$ is decreasing (resp., increasing) in c (resp., μ), being consistent with that of Ambiguity, we focus on its sensitivity to price p and variance σ^2, which exhibits a different pattern from that of Ambiguity. In particular, we also focus on the non-degenerate case in Theorem 2 that $\kappa \geq \frac{\sigma^2}{\mu^2+\sigma^2}$. In this case, recall from (1) that the optimal order quantity $q_\infty^\star$ of Ambiguity is

$$q_\infty^\star = \mu + \sigma \frac{2\kappa - 1}{2\sqrt{\kappa(1-\kappa)}} = \mu + \frac{\sigma}{2}\left(\sqrt{\frac{\kappa}{1-\kappa}} - \sqrt{\frac{1-\kappa}{\kappa}}\right). \tag{7}$$

By (7), $q_\infty^\star$ of Ambiguity is always *increasing* in p, so is $q_G^\star = G^{-1}(\kappa)$ of Nominal under any distribution G, by noting that the profit margin $\kappa = \frac{p-c}{p}$ is increasing in p. However, the monotonicity of $q_\alpha^\star$ to p can reverse that of $q_\infty^\star$ ($q_G^\star$) as shown below.

Proposition 2 (Sensitivity to Price). *There exists some $p_\alpha \geq \max\left\{\frac{\mu^2+\sigma^2}{\mu^2}c, 2\alpha\mu\right\}$ such that $q_\alpha^\star$ is decreasing in $p \in (p_\alpha, \infty)$.*

Notably, Proposition 2 points out that when price p is sufficiently large, the optimal order quantity $q_\alpha^\star$ of Misspecification, in stark contrast to $q_\infty^\star$ of Ambiguity, is *decreasing* in p; see the middle panel of Figure 5. We emphasize that such distinct sensitivity is rooted from the induced *price dependency* of the transformed worst-case demand distribution $T_{\varphi_\alpha}[G_\alpha^\star]$ in Misspecification, which stochastically reduces in the price p (see transform function φ_α in Theorem 1), while the worst-case demand distribution of Ambiguity is *independent* of the price. As the price becomes sufficiently large, the effect of reduced "demand" ($T_{\varphi_\alpha}[G_\alpha^\star]$) outweighs that of the increased profit margin, leading to the $q_\alpha^\star$ of Misspecification being decreased.

Likewise, by (7), the optimal order quantity $q_\infty^\star$ of Ambiguity is *increasing* (resp., decreasing) in σ when $\kappa \geq \frac{1}{2}$ (resp., when $\kappa < \frac{1}{2}$), under the non-degenerate condition $\kappa \geq \frac{\sigma^2}{\mu^2+\sigma^2}$, *i.e.*, $\sigma \in \left[0, \mu\sqrt{\frac{\kappa}{1-\kappa}}\right]$. Such monotonicity also holds for the optimal order quantity $q_G^\star = G^{-1}(\kappa)$ of Nominal under many commonly used distributions (*e.g.*, elliptical, uniform, and exponential). However, the situation becomes different when considering the misspecification aversion: the optimal order quantity $q_\alpha^\star$ of Misspecification, can be *decreasing* in σ when $\kappa \geq \frac{1}{2}$ in the situation of non-degeneracy; see the right panel of Figure 5. The different sensitivity pattern also uncovers an advantage of Misspecification in the solution's smoothness to parameters, by noting that the solution to Ambiguity is overly sensitive to the parameter σ as it can jump as σ changes slightly; see, *e.g.*, Embrechts et al. (2022), for various smoothness issues in optimizing risk measures.

Proposition 3 (Sensitivity to Variance). *Given $\kappa \geq \frac{1}{2}$, there exists some $\sigma_\alpha \leq \mu\sqrt{\frac{\kappa}{1-\kappa}}$ such that $q_\alpha^\star$ is decreasing in $\sigma \in \left[\sigma_\alpha, \mu\sqrt{\frac{\kappa}{1-\kappa}}\right]$.*

To summarize, sharing the mean and variance characteristics, the optimal order quantity of AMBIGUITY—hedging against the distributional uncertainty *within* the ambiguity set $\mathcal{A}$—may exhibit an *identical* sensitivity pattern to the cost-structure parameters with that of NOMINAL, and, in the situation of non-degeneracy, to the distributional characteristics with that of NOMINAL under many common distributions inside $\mathcal{A}$. For instance, both solutions of AMBIGUITY and NOMINAL share the same monotonicity to the cost c and to the mean value μ. The optimal order quantity of MISSPECIFICATION, however, hedges against another layer of distributional uncertainty *beyond* the ambiguity set $\mathcal{A}$, which therefore can *break* the sensitivity pattern of ordering characterized by the mean and variance information. This suggests that the ambiguity and misspecification, as different layers of distributional uncertainty, could result in *distinct* operational consequences, and therefore should be distinguished in the modeling.

5. Performance Guarantee

In this section, we investigate the out-of-sample performance guarantee of the optimal order quantity $q_\alpha^\star$ of MISSPECIFICATION. As we have mentioned, in many practical situations, the newsvendor has only access to incomplete knowledge on demand, and the misspecification can arise from a mixing effect of estimation error (*e.g.*, due to data limitation) and distribution shift (*e.g.*, due to non-stationarity). We consider a data-driven setting where the mean-variance ambiguity set $\mathcal{A}$ is estimated as $\mathcal{A}_N$ by using demand samples drawn from a data-generating distribution D with finite mean μ and standard deviation σ.

ASSUMPTION 1. *Random samples $\hat{v}_1, \ldots, \hat{v}_N$ are independently drawn from the data-generating distribution D, and the mean-variance ambiguity set $\mathcal{A}_N = \{G \in \mathcal{P} \mid \mathbb{E}_G[\tilde{v}] = \hat{\mu},\ \mathbb{E}_G[\tilde{v}^2] = \hat{\mu}^2 + \hat{\sigma}^2\}$ is constructed from the sample mean and variance:*

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N}\hat{v}_i > 0 \quad \text{and} \quad \hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}\hat{v}_i^2 - \left(\frac{1}{N}\sum_{i=1}^{N}\hat{v}_i\right)^2 > 0.$$

We consider the possibility that the out-of-sample distribution F can be *different* from D—a phenomenon of distribution shift. In particular, we examine the finite-sample performance guarantee through a statistical approach that provides an insightful interpretation of the performance guarantee of MISSPECIFICATION by decoupling the effects of estimation error and distribution shift.

Our analysis relies on the concentration of the estimated mean-variance ambiguity set $\mathcal{A}_N$, for which we need to investigate the optimal-transport cost $d(D, \mathcal{A}_N)$ that is closely related to the *Gelbrich distance* (Gelbrich 1990). For any $G \in \mathcal{A}_N$, the Gelbrich distance between G and the data-generating D is $\sqrt{(\hat{\mu}-\mu)^2 + (\hat{\sigma}-\sigma)^2}$, and $d(D, G) \geq (\hat{\mu}-\mu)^2 + (\hat{\sigma}-\sigma)^2$ with the inequality being tight whenever G is an affine transformation of D (as shown in Gelbrich 1990, Nguyen et al. 2021).

That is to say, if $\mathcal{A}_N$ is supported on the whole space $\mathbb{R}$, then $d(D,\mathcal{A}_N)=(\hat{\mu}-\mu)^2+(\hat{\sigma}-\sigma)^2$—the optimal-transport cost amounts to the Gelbrich distance squared. This is *not* necessarily true in our newsvendor context as the demand is non-negative and $\mathcal{A}_N$ should be supported on $\mathbb{R}_+$. Quite notably, we show that the optimal-transport cost still coincides with the Gelbrich distance squared if $\frac{\hat{\mu}}{\hat{\sigma}}\geq\frac{\mu}{\sigma}$ and otherwise, is bounded from above by the Gelbrich distance squared plus a term related to the true and estimated mean and variance—a result may be of independent interest.

Lemma 1. *Under Assumption 1, the optimal-transport cost of moving the data-generating distribution D to the mean-variance ambiguity set $\mathcal{A}_N$ can be characterized as follows.*
(i) If $\frac{\hat{\mu}}{\hat{\sigma}}\geq\frac{\mu}{\sigma}$, then $d(D,\mathcal{A}_N)=(\hat{\mu}-\mu)^2+(\hat{\sigma}-\sigma)^2$.
(ii) If $\frac{\hat{\mu}}{\hat{\sigma}}<\frac{\mu}{\sigma}$, then for sufficiently large N, it holds that

$$(\hat{\mu}-\mu)^2+(\hat{\sigma}-\sigma)^2\leq d(D,\mathcal{A}_N)\leq(\hat{\mu}-\mu)^2+(\hat{\sigma}-\sigma)^2+\frac{\mu^2\hat{\sigma}^2-\hat{\mu}^2\sigma^2}{\sigma\hat{\sigma}}.$$

We also assume the following regularity condition on the data-generating distribution D.

Assumption 2. *The data-generating distribution D is sub-Gaussian with a variance proxy ν^2, i.e., $\mathbb{E}_D[\exp(x(\tilde{v}-\mu))]\leq\exp(\frac{x^2\nu^2}{2}),\ \forall x\in\mathbb{R}$.*

Sub-Gaussianity, as a common type of light-tailed characteristics,[10] captures a wide range of distributions, including, among many others, the Gaussian distribution, the Bernoulli distribution, the uniform distribution on a convex set, and *any* bounded distributions (Vershynin 2010). With the characterization of $d(D,\mathcal{A}_N)$ in Lemma 1, we derive the following concentration inequality.

Proposition 4 (Concentration of Mean-Variance Ambiguity Set). *Under Assumptions 1 and 2, for a given confidence level $\eta\in(0,1]$, it holds for sufficiently large N that*

$$\mathbb{P}_{D^N}\left[d(D,\mathcal{A}_N)\leq\frac{(c_1+c_2\log(1/\eta))^2}{\sqrt{N}}\right]\geq 1-\eta,$$

where $c_1,c_2>0$ are constants that only depend on μ, σ, and ν.

Leveraging the concentration of ambiguity set $\mathcal{A}_N$ and the closed-form expression of the worst-case transformed expectation (Proposition 1), we can then establish a finite-sample performance guarantee of the optimal solution to Misspecification.

Theorem 3 (Finite-Sample Performance Guarantee). *Under Assumptions 1 and 2, for a given confidence level $\eta\in(0,1]$, let $\varepsilon_N=\frac{(c_1+c_2\log(1/\eta))^2}{\sqrt{N}}$ with $c_1,c_2>0$ being constants that only depend on μ, σ, and ν. Let*

$$\alpha_N=\frac{1}{2}\sqrt{\frac{p(p-c)}{\varepsilon_N+d(F,D)}},\tag{8}$$

[10] The light-tailed assumption is typically necessary for establishing the large-deviation properties for statistics of the sample mean and sample variance (Catoni 2012). As for the heavy-tailed distributions, more complex estimation procedures (for the mean and variance) are needed to achieve acceptable convergence rates (Cai et al. 2010).

if $\varepsilon_N + d(F, D) < \kappa(\hat{\mu} - \hat{\sigma}\sqrt{(1-\kappa)/\kappa})^2$*; and* $\alpha_N = 0$*, otherwise. Consider the optimal solution* $q^\star_{\alpha_N}$ *and the optimal value* $\Pi^\star_{\alpha_N}$ *of* Misspecification *with* $\alpha = \alpha_N$ *and* $\mathcal{A} = \mathcal{A}_N$*. For sufficiently large* N*, it holds that*

$$\mathbb{P}_{D^N}\left[\mathbb{E}_F[\pi(q^\star_{\alpha_N}, \tilde{u})] \geq \Bigg(\underbrace{\Pi^\star_{\alpha_N}}_{\substack{\text{in-sample}\\ \text{optimal value}}} - \frac{1}{2}\sqrt{\underbrace{p(p-c)\varepsilon_N}_{\substack{\text{effect of}\\ \text{estimation error}}} + \underbrace{p(p-c)d(F,D)}_{\substack{\text{effect of}\\ \text{distribution shift}}}}\Bigg)^+\right] \geq 1 - \eta,$$

where D^N *is the* N*-fold product of* D*.*

The guarantee derived in Theorem 3 is the in-sample optimal value of Misspecification subtracting the out-of-sample misspecification effect described by the estimation error in mean and variance and the distribution shift. For the estimation error, it is related to the upper bound ε_N on the distance $d(D, \mathcal{A}_N)$ between the data-generating distribution and the estimated mean-variance ambiguity set, which diminishes as $N \to \infty$. For the distribution shift captured by $d(F, D)$, it is *independent* of the sample size N. A key statistical implication is that as long as the out-of-sample distribution F shifts from the data-generating distribution D, there is always a constant amount of loss $\frac{1}{2}\sqrt{p(p-c)d(F,D)}$ in terms of the performance guarantee, even as $N \to \infty$ ($\varepsilon_N \to 0$).

Theorem 3 also suggests that the calibration for the index α_N of misspecification aversion is affected by both the estimation error and the distribution shift. According to (8), when non-zero, α_N is increasing in the sample size N (*i.e.*, decreasing in ε_N) while decreasing in the extent of distribution shift $d(F, D)$. This implies that at the same confidence level, the newsvendor needs to be more misspecification averse in either case of a smaller amount of data or a more significant distribution shift, to guarantee the performance. Moreover, in the presence of distribution shift ($d(F, D) > 0$), even when the estimation error vanishes with the sufficient data (*i.e.*, $\varepsilon_N \to 0$ as $N \to \infty$), $\alpha_N \to \frac{1}{2}\sqrt{\frac{p(p-c)}{d(F,D)}} < \infty$, implying that Misspecification ($\alpha_N < \infty$) shall still outperform Ambiguity ($\alpha_N = \infty$). Importantly, this means that the commonly used cross-validation approach (which is based on the data-generating distribution D) for calibrating distributionally robust optimization models could, unfortunately, work *poorly* in the situation of misspecification (*i.e.*, calibrating the index α for Misspecification), as we demonstrate in the numerical study.

6. Extensions

In this section, we extend the model Misspecification along the following directions: (*i*) there are multiple products, (*ii*) the ambiguity set is defined via optimal transport, and (*iii*) the extent of misspecification is measured by total variation distance.

6.1. Multiple Products

Our misspecification-averse model can be extended to the multi-product newsvendor problem. Consider M products (each with unit price p_i and cost c_i, $i \in [M]$) whose random demands are collectively denoted by $\tilde{\boldsymbol{u}} = (\tilde{u}_1, \ldots, \tilde{u}_M)$ that follows a multi-dimensional distribution F. The misspecification-averse newsvendor then solves

$$\max_{\boldsymbol{q}\geq \boldsymbol{0}} \min_{F\in\mathcal{P}_M} \{\mathbb{E}_F[\omega(\boldsymbol{q}, \tilde{\boldsymbol{u}})] + \alpha \cdot d(F, \mathcal{C})\}, \qquad \text{(MULTIPLE)}$$

where $\mathcal{P}_M$ is the set of probability distributions supported on $\mathbb{R}^M_+$, the optimal-transport cost $d(\cdot,\cdot)$ is defined in (2) with the cost function $\|\cdot\|_2^2$, $\boldsymbol{q} = (q_1, \ldots, q_M)$ is the vector of order quantities, and $\omega(\boldsymbol{q}, \boldsymbol{u}) = \sum_{i=1}^M \pi_i(q_i, u_i) = \sum_{i=1}^M p_i \cdot \min\{q_i, u_i\} - c_i q_i$.

If the ambiguity set $\mathcal{C}$ is specified by marginal mean-variance information of each product, then MULTIPLE is separable concerning products and Theorem 2 yields each product's optimal order quantity. If $\mathcal{C}$ is specified by mean and correlation information, then MULTIPLE becomes much more involved, as its ambiguity-averse counterpart is already intractable (Hanasusanto et al. 2015, Natarajan et al. 2018). More details on the reformulation and computational difficulty of MULTIPLE with complete covariance information are presented in Section EC.6. In the following, we show that an analytical solution can be derived for a case that captures the partial correlation across products. In particular, we consider an ambiguity set with mean and sum-of-variance constraints:

$$\mathcal{C} = \left\{ G \in \mathcal{P}_M \;\middle|\; \mathbb{E}_G[\tilde{v}_i] = \mu_i \;\forall i \in [M],\; \mathbb{E}_G\left[\sum_{i\in[M]} \tilde{v}_i^2\right] \leq K \right\},$$

where K is some non-negative constant that bounds the sum of the variance of products' demands. Note that when $M = 1$ (*i.e.*, there is a single product), the ambiguity set $\mathcal{C}$ reduces to the mean-variance ambiguity set $\mathcal{A}$. The analytical solution to MULTIPLE with $\mathcal{C}$ relies on characterizing the optimal dual variable, denoted by $\lambda^\star$, to its sum-of-variance constraint. We indicate that adopting a similar technique, we can also obtain the optimal solution of MULTIPLE under an additional budget constraint $\sum_{i\in[M]} q_i \leq Q$ for some $Q > 0$.

To ease our presentation, let $\bar{\lambda}_0 = 0$, $\bar{\lambda}_i = \frac{c_i}{(2\mu_i - p_i/\alpha)^+}$ for $i \in [M]$ and $\bar{\lambda}_{M+1} = +\infty$. Without loss of generality, we rearrange $\bar{\lambda}_i, i \in [M]$ ascendingly, *i.e.*, $\bar{\lambda}_1 \leq \bar{\lambda}_2 \leq \ldots \leq \bar{\lambda}_M$. Besides, we define

$$i^\star = \min\{j \in [M+1] \mid \Theta_j(\bar{\lambda}_j) < K\}, \qquad (9)$$

where for $j \in [M+1]$ and $\lambda \geq 0$,

$$\Theta_j(\lambda) = \sum_{i\in[M]\setminus[j-1]} \frac{p_i \mu_i^2(\alpha^2 c_i + 2\alpha c_i \lambda + p_i \lambda^2)}{(p_i\lambda + \alpha c_i)^2} + \sum_{i\in[j-1]} \frac{(p_i - c_i)c_i}{4\lambda^2}. \qquad (10)$$

We then present the following strong-duality result that characterizes the optimal dual variable to the sum-of-variance constraint and the *decomposability* of the dual problem.

THEOREM 4. *Given $K \geq 0$ and $\alpha \geq 0$,* MULTIPLE *is equivalent to*

$$\max_{\lambda \geq 0} \left\{ -\lambda K + \sum_{i \in [M]} \Pi^{\star}_{i,\alpha}(\lambda) \right\}.$$

Here, for each $i \in [M]$, $\Pi^{\star}_{i,\alpha}(\lambda)$ is the optimal value of the following optimization problem:

$$\Pi^{\star}_{i,\alpha}(\lambda) = \max_{q_i \geq 0} \min_{F_i \in \mathcal{P}} \left\{ \mathbb{E}_{F_i}[\pi_i(q_i, \tilde{u}_i)] + \min_{G_i \in \mathcal{C}_i} \left\{ \lambda \cdot \mathbb{E}_{G_i}[\tilde{v}_i^2] + \alpha \cdot d(F_i, G_i) \right\} \right\} \tag{11}$$

with $\tilde{u}_i \sim F_i$, $\tilde{v}_i \sim G_i$, and $\mathcal{C}_i = \{G \in \mathcal{P} \mid \mathbb{E}_G[\tilde{v}] = \mu_i\}$ being a mean ambiguity set. The optimal $\lambda^\star$ is decreasing in K, and with $i^\star$ in (9) and $\Theta_i(\cdot)$ in (10), it can be characterized as follows.

(i) If $\Theta_{i^\star}(\bar{\lambda}_{i^\star-1}) \leq K$, then $\lambda^\star = \bar{\lambda}_{i^\star-1}$.

(ii) If $\Theta_{i^\star}(\bar{\lambda}_{i^\star-1}) > K$, then $\lambda^\star \in (\bar{\lambda}_{i^\star-1}, \bar{\lambda}_{i^\star})$ is the solution to the equation $\Theta_{i^\star}(\lambda) = K$.

By Theorem 4, the value of $\lambda^\star$ can be efficiently binary searched due to monotonicity. Given the value of $\lambda^\star$, MULTIPLE can be decomposed into multiple misspecification-averse *single-product* newsvendor problems in (11). In particular, each single-product problem is now regularized by

$$\min_{G_i \in \mathcal{C}_i} \{\lambda^\star \cdot \mathbb{E}_{G_i}[\tilde{v}_i^2] + \alpha \cdot d(F_i, G_i)\}.$$

Moreover, it is critical to note that the optimal dual variable $\lambda^\star$ captures not only the demand-correlation information encoded in the sum-of-variance constraint, but also the cross-product cost structure (*i.e.*, p_i and c_i, $i \in [M]$) that collectively affects those single-product problems. The following result adapts the reasoning for Theorem 2 to derive the analytical solution for MULTIPLE.

THEOREM 5 (OPTIMAL SOLUTION: MULTIPLE PRODUCTS). *Given $\alpha \geq 0$ and the optimal $\lambda^\star$ characterized in Theorem 4, the optimal order quantity $\boldsymbol{q}^\star_\alpha$ of* MULTIPLE *is, for each $i \in [M]$, $q^\star_{i,\alpha} = \mu_i + \frac{p_i - 2c_i}{4\lambda^\star} - \frac{p_i}{4\alpha}$ if $\alpha \geq \frac{p_i}{2(\mu_i - c_i/(2\lambda^\star))^+}$ and $q^\star_{i,\alpha} = \frac{\lambda^\star(\lambda^\star+\alpha)p_i\mu_i^2}{\alpha(p_i\lambda^\star/\alpha + c_i)^2}$ otherwise.*

The analytical solution $\boldsymbol{q}^\star_\alpha$ captures the information of distribution characteristics, cost structure, and correlation across products. In particular, the optimal order quantity $q^\star_{i,\alpha}$ for each product i is not only determined by its distribution characteristic (*i.e.*, μ_i) and cost structure (*i.e*, p_i and c_i), but also by those of other products via $\lambda^\star$. Furthermore, in the multi-product problem, the index of misspecification aversion α has a *double effect* on the optimal order quantity in two ways: on the one hand, it directly affects the optimal order quantity $q^\star_{i,\alpha}$ of each product in the decomposed single-product problem (11); on the other hand, it also affects the optimal dual variable $\lambda^\star$ that in turn influences the optimal order quantities $q^\star_{i,\alpha}$ of all products.

Finally, we emphasize that Theorem 5 generalizes Theorem 2 from a single product to multiple products (see Section EC.5 for a detailed derivation) as well as the forthcoming Corollary 1 for the ambiguity-averse multi-product model to ambiguity and misspecification aversion. Note that as $\alpha \to \infty$, MULTIPLE immediately reduces to the ambiguity-averse only counterpart

$$\max_{\boldsymbol{q}\geq \boldsymbol{0}} \min_{G\in\mathcal{C}} \mathbb{E}_G[\omega(\boldsymbol{q},\tilde{\boldsymbol{v}})]. \tag{12}$$

Its optimal solution, as expected, generalizes the Scarf model from a single product to multiple products. The proof is straightforward and is thus omitted.

COROLLARY 1. *Let the optimal dual variable $\lambda^\star$ be characterized in Theorem 4, the optimal quantity $\boldsymbol{q}^\star_\infty$ of the ambiguity-averse multi-product newsvendor problem (12) is, for each $i\in[M]$, $q^\star_{i,\infty} = \mu_i + \frac{p_i - c_i}{4\lambda^\star}$ if $\lambda^\star \geq \frac{c_i}{2\mu_i}$ and $q^\star_{i,\infty} = \frac{p_i\mu_i^2}{c_i^2}\cdot\lambda^\star$ otherwise.*

6.2. Distance-Based Ambiguity Set

Apart from the mean-variance ambiguity set $\mathcal{A}$, the following distance-based ambiguity set

$$\mathcal{B}(\theta) = \{G\in\mathcal{P} \mid d(G,H)\leq\theta\}$$

with a reference distribution H, $\theta\geq 0$, and optimal-transport cost $d(\cdot,\cdot)$ between probability distributions is a popular alternative for specifying partial distributional information, which has also been widely used in aforementioned applications of decision theory (Petracou et al. 2022), newsvendor (Chen and Xie 2021, Zhang et al. 2023), and risk management (Wozabal 2014). Conceptually, $\mathcal{B}(\theta)$ consists of all probability distributions in a θ-neighbourhood around H, where the closeness is measured by $d(\cdot,\cdot)$ given in (2). Since $\mathcal{B}(\theta_1)\subseteq\mathcal{B}(\theta_2)$ for any $\theta_2\geq\theta_1\geq 0$, a larger value of θ indicates a lower confidence in H. When $\theta=0$, $\mathcal{B}(\theta)$ shrinks to a singleton containing only the reference distribution H, that is, $\mathcal{B}(0)=\{H\}$. It is natural to consider the following variant of MISSPECIFICATION where we replace $\mathcal{A}$ with $\mathcal{B}(\theta)$:

$$\max_{q\geq 0}\min_{F\in\mathcal{P}} \left\{\mathbb{E}_F[\pi(q,\tilde{u})] + \alpha\cdot d(F,\mathcal{B}(\theta))\right\}, \tag{13}$$

which hedges against the possible misspecification over the ambiguity set $\mathcal{B}(\theta)$. Quite notably, problem (13) is essentially equivalent to hedging against misspecification over the singleton $\{H\}$ but with a *stronger* aversion to misspecification. To avoid a degenerate case, we assume $H^{-1}(\kappa)>0$.

THEOREM 6 (OPTIMAL SOLUTION: DISTANCE-BASED AMBIGUITY SET). *Given $\theta\geq 0$, $\alpha\geq 0$ and a reference distribution H, there exists some $\gamma^\star\in[0,\alpha]$ such that problem (13) is equivalent to*

$$\max_{\psi\geq 0}\min_{F\in\mathcal{P}} \left\{\mathbb{E}_F[\pi(\psi,\tilde{u})] + \gamma^\star\cdot d(F,H)\right\}. \tag{14}$$

When $\theta=0$, $\gamma^\star=\alpha$; otherwise, with the optimal order quantity $q^\star_H = H^{-1}(\kappa)$ of NOMINAL *under H and $\beta=\int_0^{q^\star_H} u^2\mathrm{d}H(u)>0$, $\gamma^\star$ can be characterized as follows.*

(i) *If* $\theta \geq \beta$, *then* $\gamma^\star = 0$.

(ii) *If* $\theta < \beta$ *and* $\alpha(1-\sqrt{\theta/\beta}) < \frac{p}{2q_H^\star}$, *then* $\gamma^\star = \alpha(1-\sqrt{\theta/\beta})$.

(iii) *If* $\theta < \beta$ *and* $\alpha(1-\sqrt{\theta/\beta}) \geq \frac{p}{2q_H^\star}$, *then* $\gamma^\star$ *is the solution to*

$$\int_0^{\frac{p}{2x}} u^2 \mathrm{d}H(u) + \frac{p^2}{4x^2}\left(\kappa - H\left(\frac{p}{2x}\right)\right) - \frac{\alpha^2\theta}{(\alpha - x)^2} = 0.$$

With $\gamma^\star$, *the optimal order quantity* $\psi^\star_{\gamma^\star}$ *of problem* (14) *can be characterized as* $\psi^\star_{\gamma^\star} = q_H^\star \cdot (\frac{\gamma^\star q_H^\star}{p})$ *if* $\gamma^\star < \frac{p}{2q_H^\star}$ *and* $\psi^\star_{\gamma^\star} = q_H^\star - \frac{p}{4\gamma^\star}$ *otherwise.*

Theorem 6 establishes the equivalence between problem (13)—which hedges against misspecification over a distance-based ambiguity set $\mathcal{B}(\theta)$ around the reference distribution—and problem (14) that, with a stronger aversion, hedges against misspecification to the reference distribution H. Note that Theorem 6 states that $\gamma^\star = \alpha$ whenever $\theta = 0$—which, indeed, corresponds to the ambiguity neutrality. If we further have $\alpha \to \infty$, then problem (13) reduces to Nominal under H, and as expected, Theorem 6 concludes that $\psi^\star_{\gamma^\star} = q_H^\star$.

6.3. Misspecification Measured by Total Variation Distance

Apart from the optimal-transport cost, total variation distance is also popular for measuring the closeness between probability distributions. In this section, we replace $d(\cdot,\cdot)$ in (2) with total variation distance when defining $d(F,\mathcal{A})$ and investigate the corresponding Misspecification problem. Formally, the total variation distance between distributions F and G is defined as

$$d_{\mathrm{TV}}(F,G) = \sup_{A \in \mathcal{B}(\mathbb{R}_+)} \left|F(A) - G(A)\right|,$$

where $\mathcal{B}(\mathbb{R}_+)$ is the Borel σ-algebra on $\mathbb{R}_+$. Equipped with the total variation distance,[11] we investigate the following variant of Misspecification:

$$\max_{q \geq 0} \min_{G \in \mathcal{A}} \min_{F \in \mathcal{P}} \{\mathbb{E}_F[\pi(q,\tilde{v})] + \alpha \cdot d_{\mathrm{TV}}(F,G)\}. \tag{15}$$

Theorem 7 (Optimal Solution: TV-Based Misspecification). *Given* $\alpha \geq 0$ *and the mean-variance ambiguity set* $\mathcal{A}$, *problem* (15) *can be equivalently reformulated as*

$$\max_{0 \leq q \leq \frac{2\alpha}{p}} \min_{G \in \mathcal{A}} \mathbb{E}_G[\pi(q,\tilde{v})],$$

and its optimal order quantity is $q_\alpha^\star = \min\{\frac{2\alpha}{p}, q_\infty^\star\}$, *where* $q_\infty^\star$ *is the optimal order quantity of* Ambiguity *characterized in* (1).

[11] Other types of ϕ-divergence can also be applied to problem (15) of misspecification, which however are less computationally appealing. For instance, if $d(\cdot,\cdot)$ is the Kullback–Leibler divergence, then problem (15) becomes optimizing the worst-case constant absolute risk aversion, which is generally intractable (Chen and Sim 2025). Also, if we take $d(\cdot,\cdot)$ as the Gini concentration index, then problem (15) becomes optimizing worst-case mean-variance, which does not admit analytical solutions either.

Theorem 7 reveals that the optimal order quantity $q^\star_\alpha$ of problem (15) grows increasingly and linearly in α, capped by the optimal order quantity $q^\star_\infty$ of AMBIGUITY. When α is sufficiently large (that is, $\alpha \geq \frac{p}{2}q^\star_\infty$), $q^\star_\alpha$ coincides with $q^\star_\infty$. In this case, the newsvendor fully trusts the information specified in $\mathcal{A}$ and essentially becomes misspecification neutral to the ambiguity set $\mathcal{A}$. Hence, ambiguity aversion takes full charge of determining the formula of $q^\star_\alpha$. When α is relatively small (that is, $\alpha < \frac{p}{2}q^\star_\infty$), $q^\star_\alpha = \frac{2\alpha}{p}$ is then purely determined by misspecification aversion without being affected by the mean-variance information in $\mathcal{A}$.

7. Numerical Experiments with Retailing Data

In this section, we demonstrate the effectiveness of incorporating misspecification, using the real-world daily demand data over one year for different stock keeping units (SKUs) of our industrial partner (a supermarket). Our goal is to compare the out-of-sample expected profit of the optimal order quantities obtained from MISSPECIFICATION and AMBIGUITY, as well as NOMINAL.

7.1. Data and Models

Our data consists of a set of SKUs, and we first consider the popular drinking water (as mentioned in Figure 1 and denoted by SKU_0) that shows non-stationary characteristics, especially in the sense of monthly mean and variance. For the experimental purpose, we consider the demand data in two consecutive months as training and testing samples. In particular, according to the variability in sales series and the associated mean-variance change over the consecutive months as demonstrated in Figure 1, we identify October vs. November as a *low-variability* scenario, January vs. February as a *moderate-variability* scenario, and August vs. September as a *high-variability* scenario. See the first column of Figure 6, highlighting the non-stationarity of the demand.

At the end of the training month, the newsvendor obtains the demand observations (training samples) $\hat{v}_1, \ldots, \hat{v}_N$ and needs to decide an order quantity to satisfy the random demand that will materialize in the testing month. The newsvendor solves NOMINAL with the empirical distribution based on $\hat{v}_1, \ldots, \hat{v}_N$. For AMBIGUITY, the newsvendor estimates μ and σ^2 of the mean-variance ambiguity set $\mathcal{A}$ via sample mean and sample variance, that is, $\mu = \frac{1}{N}\sum_{i=1}^N \hat{v}_i$ and $\sigma^2 = \frac{1}{N}\sum_{i=1}^N \hat{v}_i^2 - \left(\frac{1}{N}\sum_{i=1}^N \hat{v}_i\right)^2$. However, taking these estimates as the mean and variance of the demand in the testing month may lead to misspecification (recall from Figure 1). It is thus meaningful to consider MISSPECIFICATION with different values of α. For a comprehensive comparison, we also consider two other benchmarks for AMBIGUITY. The first is the Wasserstein newsvendor model, *i.e.*,

$$\max_{q\geq 0} \min_{G\in\mathcal{B}(\theta)} \mathbb{E}_G[\pi(q,\tilde{v})], \tag{WASSERSTEIN}$$

where $\mathcal{B}(\theta)$ is defined in Section 6.2 and its solution can be derived via Theorem 6 with $\alpha \to \infty$. The other is the newsvendor model based on the ambiguity set proposed in Delage and Ye (2010):

$$\max_{q\geq 0} \min_{G\in\mathcal{D}} \mathbb{E}_G[\pi(q,\tilde{v})], \qquad \text{(Delage-Ye)}$$

where $\mathcal{D} = \{G \in \mathcal{P} \mid \mu - \sqrt{\gamma_1}\sigma \leq \mathbb{E}_G[\tilde{v}] \leq \mu + \sqrt{\gamma_1}\sigma,\ \mathbb{E}_G[\tilde{v}^2] \leq -\mu^2 + \gamma_2\sigma^2 + 2\mu\mathbb{E}_G[\tilde{v}]\}$, and $\gamma_1 \geq 0$ and $\gamma_2 \geq 1$ are two constants. Delage-Ye does not admit any closed-form solution and is solved using the reformulation in Delage and Ye (2010). Hyperparameters θ and (γ_1, γ_2) in these two benchmarks are selected via the cross-validation approach. In fact, we also test the performance of the newsvendor model with uncertain mean and variance as depicted in Remark 1, which turns out to be very similar to that of Delage-Ye.

7.2. Results and Discussion

Figure 6 summarizes the results of different approaches in the three scenarios where the variability between training and testing samples is, respectively, low, moderate, and high. Regarding the in-sample expected profit, in all scenarios, Nominal achieves the highest value, while Misspecification converges to Ambiguity as α approaches infinity. Focusing on the out-of-sample performance, *first*, in the low-variability scenario, Nominal outperforms Ambiguity, and under most values of α, Misspecification does not yield a higher out-of-sample expected profit than Nominal—an intuitive result since the demand process is quite stationary. For Delage-Ye and Wasserstein, they outperform Ambiguity and Nominal, and almost achieve the best performance of Misspecification. *Furthermore*, in the moderate-variability scenario, Nominal yields \$3.19 in the out-of-sample expected profit while Misspecification (resp., Ambiguity) results in an improvement at most \$0.54 or 16.93% in percentage (resp., an improvement of \$0.25 or 7.84%). We also note that Delage-Ye achieves the same expected profit as Ambiguity while Wasserstein yields an improvement of \$0.09 or 2.68% compared to Nominal. *Finally*, in the high-variability scenario, Ambiguity and Misspecification have an even larger improvement in out-of-sample expected profit: Nominal yields \$3.24 while Misspecification (resp., Ambiguity) results in an improvement at most \$2.65 or 81.79% (resp., an improvement of \$1.18 or 36.42%). We observe that Delage-Ye also achieves the same expected profit as Ambiguity while Wasserstein yields an improvement of \$0.10 or 3.06% compared to Nominal.

In the third column of Figure 6, we first emphasize two values of the index of misspecification aversion, α^{CV} and $\alpha^{\star}$. The former is selected via cross-validation using the training data, while the latter is the one that achieves the largest out-of-sample expected profit. We would like to highlight two important observations and insights as follows.

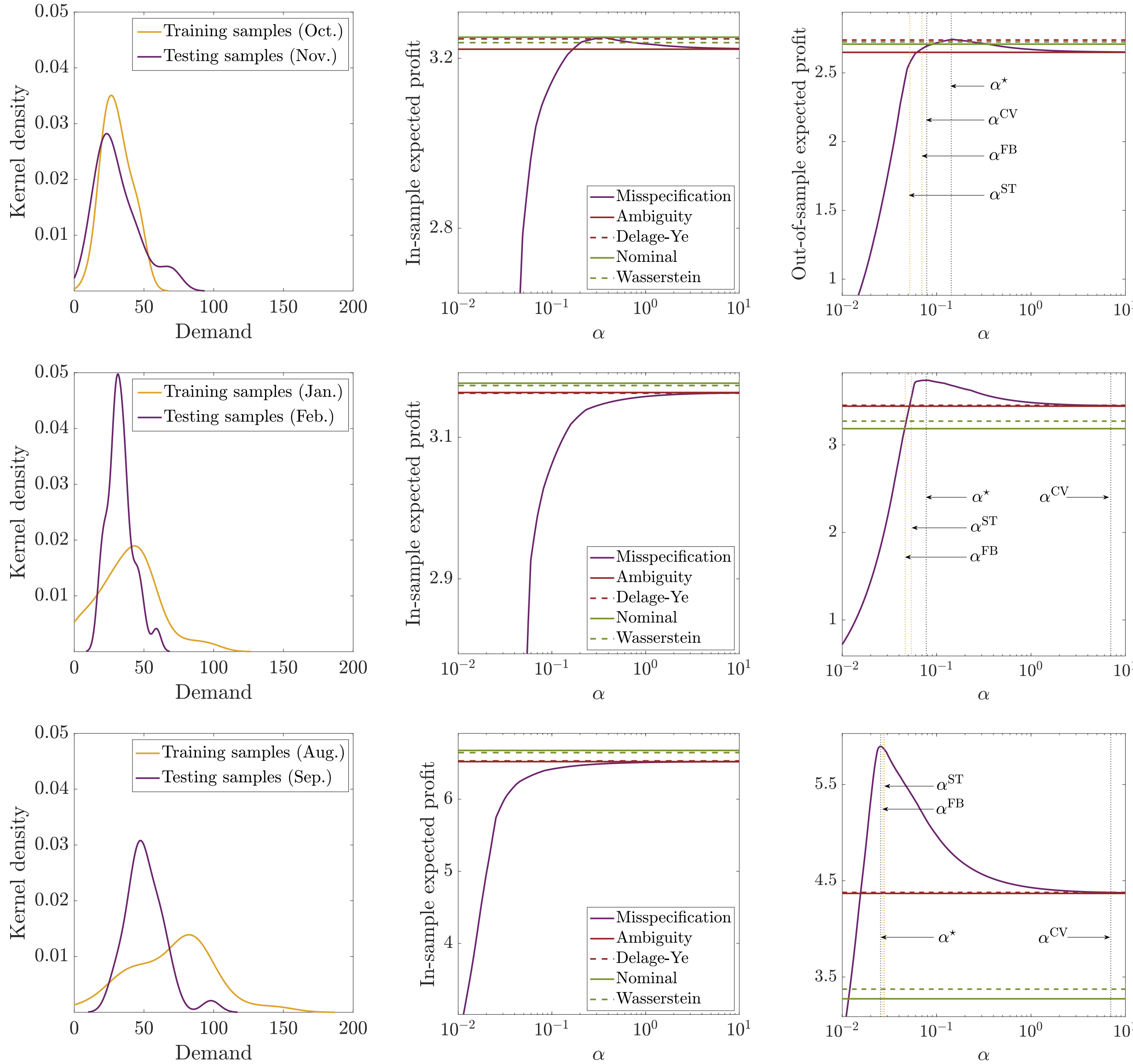


Figure 6 Demand density and performance of MISSPECIFICATION, AMBIGUITY, DELAGE-YE, NOMINAL, and WASSERSTEIN when variability is low (first row), moderate (second row), and high (third row).

1. In the low-variability scenario with stationary demand, MISSPECIFICATION with α^{CV} performs quite close to NOMINAL and AMBIGUITY, and α^{CV} is also close to $\alpha^\star$. This, not only justifies the predominance of mean-variance statistics in capturing the underlying distribution for the newsvendor's decision, but also implies the usefulness of the cross-validation method in calibrating the parameter α of MISSPECIFICATION, in the situation of stationary demand.
2. In the moderate-variability and high-variability scenarios with non-stationary demand, the calibrated MISSPECIFICATION model with $\alpha = \alpha^{\mathrm{CV}}$ yields an out-of-sample performance close to that of AMBIGUITY, which, however, is *far away* from the best performance that MISSPECIFICATION could achieve with $\alpha = \alpha^\star$. This confirms the implication of the finite-sample

performance guarantee derived in Theorem 3: in the situation of distribution shift where the testing samples vary highly from the training samples, cross-validation—purely relying on training samples—could be depreciative in its effectiveness for calibrating a model's parameter.

It is generally impossible to calibrate the index α of misspecification aversion under arbitrary and unknown distribution shifts (Sutter et al. 2021). However, in certain practical situations (for example, companies may conduct test markets before launching a new product to gather limited testing data; see Urban and Hauser 1993), partial information about the distribution shift may be available through a limited number of testing samples drawn from the out-of-sample distribution (Sugiyama and Kawanabe 2012). To address such cases, we propose two calibration strategies: a formula-based approach and a stress-testing approach. We provide a brief description in the following and leave the implementation details to Section EC.7.

The formula-based approach directly explores formula (8) derived in Theorem 3 to determine α. Specifically, we use testing samples to estimate $d(F, D)$, while treating ϵ_N as a tuning parameter and using cross-validation to select ϵ_N.

The stress-testing approach can be viewed as a controlled stress test that proceeds with two steps: constructing a stress-testing distribution and validating based on the constructed stress-testing distribution. In the construction step, to mimic the testing environment, we build a synthetic stress-testing distribution that matches the estimated distribution shift based on the testing samples. In the validation step, we cross-validate α based on the constructed testing distribution.

The third column of Figure 6 shows that in all three testing scenarios, both α^{FB} (by the formula-based approach) and α^{ST} (by the stress-testing approach) select relatively small values that, compared to α^{CV}, are more closer to the optimal $\alpha^{\star}$.

1. In the low-variability scenario—where the distribution shift is minimal—the formula-based α^{FB} closely aligns with α^{CV} (which is already good enough), while the stress-testing α^{ST} tends to yield smaller values. This behavior is expected: although the actual shift is small, the empirical distance between the training distribution and the few testing samples often overestimates the shift. As a result, both α^{ST} and α^{FB} are calibrated to be relatively conservative.
2. In the moderate-variability scenario—where the distribution shift is more pronounced—the empirical distance provides a more accurate reflection of the shift compared to the low-variability scenario. We observe that the stress-testing approach yields an α^{ST} with slightly better out-of-sample performance than α^{CV}, while the formula-based α^{FB} performs marginally worse. Notably, both α^{FB} and α^{ST} are closer to the optimal $\alpha^{\star}$ than in the low-variability scenario. When the distribution shift is substantial, the empirical distance reliably captures the extent of the shift. Consequently, both α^{FB} and α^{ST} yield better-calibrated estimates that closely approximate the optimal α, and they outperform the plain vanilla α^{CV}, which is validated solely over training samples.

α	Performance Comparison	Percentage	Π_{M} Mean | STD	Π_{A} Mean | STD	Π_{S} Mean | STD
α_{low}	$\Pi_{\mathrm{M}} > \Pi_{\mathrm{S}}$ and $\Pi_{\mathrm{M}} > \Pi_{\mathrm{A}}$	28%	24.95 | 19.93	14.49 | 13.17	15.90 | 14.33
	$\Pi_{\mathrm{M}} \le \Pi_{\mathrm{S}}$ or $\Pi_{\mathrm{M}} \le \Pi_{\mathrm{A}}$	72%	5.09 | 7.22	8.60 | 8.94	8.80 | 8.45
α_{mid}	$\Pi_{\mathrm{M}} > \Pi_{\mathrm{S}}$ and $\Pi_{\mathrm{M}} > \Pi_{\mathrm{A}}$	81%	12.07 | 10.95	9.74 | 9.72	9.56 | 9.15
	$\Pi_{\mathrm{M}} \le \Pi_{\mathrm{S}}$ or $\Pi_{\mathrm{M}} \le \Pi_{\mathrm{A}}$	19%	13.28 | 14.55	12.52 | 13.59	15.91 | 15.51
α_{high}	$\Pi_{\mathrm{M}} > \Pi_{\mathrm{S}}$ and $\Pi_{\mathrm{M}} > \Pi_{\mathrm{A}}$	69%	11.49 | 11.04	10.03 | 10.10	9.58 | 9.50
	$\Pi_{\mathrm{M}} \le \Pi_{\mathrm{S}}$ or $\Pi_{\mathrm{M}} \le \Pi_{\mathrm{A}}$	31%	11.47 | 11.55	10.69 | 11.39	13.06 | 12.84

Table 1 Out-of-sample expected profits over 100 SKUs. For each SKU, $\alpha_{\text{low}} = p/100$, $\alpha_{\text{mid}} = p/20$, and $\alpha_{\text{high}} = p/10$, where p is the unit price of the SKU. Here, we denote by Π_{M}, Π_{A}, and Π_{S} the out-of-sample expected profits of Misspecification, Ambiguity, and Nominal, respectively.

These findings demonstrate that with limited knowledge of the distribution shift, our proposed calibration strategies perform effectively in mitigating its impact—particularly in scenarios with substantial variability. We would like to acknowledge that our proposal is merely an initial attempt, and that developing a rigorous statistical framework for inferring distributional shifts remains an active and emerging area of research. As such, this aspect lies beyond the scope of the present study and is deferred to future work, as discussed in the conclusion.

We next repeat the above experiment over a pool of 100 SKUs, for each of which we randomly select two consecutive months as training and testing samples. For each SKU, we consider the same setting as in Section 7.1 to evaluate the out-of-sample expected profits of Misspecification, Ambiguity, and Nominal. Table 1 summarizes the number of SKUs that one model outperforms another, and the corresponding mean and standard deviation of the out-of-sample profits of a model over these SKUs. Under a small value of α (*i.e.*, α_{low}), for 28% of 100 SKUs, Misspecification outperforms both Ambiguity and Nominal with a large profit improvement but a large standard deviation; for the majority 72%, Misspecification underperforms either Ambiguity or Nominal with a large profit loss and a small standard deviation. Under a medium value of α (*i.e.*, α_{mid}), Misspecification yields superior performance than both Ambiguity and Nominal by noting that Misspecification outperforms both Ambiguity and Nominal for a majority 81% of 100 SKUs. Even under a high value of α (*i.e.*, α_{high}), Misspecification also has a fairly good out-of-sample performance, where Misspecification outperforms both Ambiguity and Nominal for a majority 69% of 100 SKUs. In other words, for each $\alpha \in \{\alpha_{\text{low}}, \alpha_{\text{mid}}, \alpha_{\text{high}}\}$, there always quite a proportion of SKUs such that over these products Misspecification has a better out-of-sample performance than both Ambiguity and Nominal, justifying the need of incorporating misspecification to the newsvendor problem.

8. Conclusion

Since the seminal work of Scarf (1958), the mean-variance ambiguity set has been widely used for decision-making in mitigating distributional uncertainty. However, in many practical situations, the mean and variance can be misspecified, consequently resulting in inexpert newsvendor decisions. To address this issue, we introduce misspecification upon ambiguity and propose a misspecification-averse (and ambiguity-averse) newsvendor model. We investigate the impact and rationale of misspecification aversion from decision-criterion, operational, and statistical perspectives. We also extend our model to establish a comprehensive framework (multi-products, ambiguity captured by optimal transport, and misspecification measured by total variation distance) for the newsvendor under ambiguity and misspecification.

Our present study focuses on and investigates many aspects of the misspecification-averse newsvendor problem. The framework has several interesting directions remaining unexplored and can be extended to other operational problems, opening up promising avenues for future studies.

Statistical inference of distribution shift. As articulated in Theorem 3, the performance guarantee and the index of misspecification aversion are statistically described with a term of distribution shift. How to estimate the distribution shift is statistically critical and is also practically relevant for calibrating the model of misspecification aversion.

Misspecification in prescriptive analytics. Prescriptive analytics, as an emerging paradigm for data-driven decision-making, seeks a decision rule that maps the observed data to an action, which usually leverages some parametric (structural) assumptions on the uncertainty (see, *e.g.*, Qi and Shen 2022, Chu et al. 2025). These assumed parametric models could misspecify the ground truth. Therefore, we believe that our approach, in marriage with the prescriptive analytics framework, has the potential to mitigate the downside consequences of model misspecification.

Structuring distributional uncertainty in other operations management problems. Although this study focuses on the newsvendor problem, our analysis can also be applied in other operations management settings, for instance, inventory control, logistics, dynamic pricing, and project management, wherein the distributional uncertainty has been widely acknowledged. Following the spirit of "all models are wrong, but some are useful" (Box 1976), we believe that treating ambiguity and misspecification differently in a properly structured fashion can differentiate the *useful* wrong models and *harmful* wrong models in coping with uncertainty, and therefore enhance the decisions for operations management.

References

An, Lin, Andrew Li, Benjamin Moseley, R Ravi. 2025. The nonstationary newsvendor with (and without) predictions. *Forthcoming in Manufacturing & Service Operations Management.*

Ardestani-Jaafari, Amir, Erick Delage. 2016. Robust optimization of sums of piecewise linear functions with application to inventory problems. *Operations Research* **64**(2) 474–494.

Arrow, Kenneth, Theodore Harris, Jacob Marschak. 1951. Optimal inventory policy. *Econometrica* **19**(3) 250–272.

Aydogan, Ilke, Loïc Berger, Valentina Bosetti, Ning Liu. 2023. Three layers of uncertainty. *Journal of the European Economic Association* **21**(5) 2209–2236.

Ben-Tal, Aharon, Stephen Boyd, Arkadi Nemirovski. 2006. Extending scope of robust optimization: Comprehensive robust counterparts of uncertain problems. *Mathematical Programming* **107**(1-2) 63–89.

Ben-Tal, Aharon, Ruud Brekelmans, Dick den Hertog, Jean-Philippe Vial. 2017. Globalized robust optimization for nonlinear uncertain inequalities. *INFORMS Journal on Computing* **29**(2) 350–366.

Besbes, Omar, Dan Iancu, Nikolaos Trichakis. 2018. Dynamic pricing under debt: Spiraling distortions and efficiency losses. *Management Science* **64**(10) 4572–4589.

Box, George. 1976. Science and statistics. *Journal of the American Statistical Association* **71**(356) 791–799.

Cai, Tony, Cun-Hui Zhang, Harrison Zhou. 2010. Optimal rates of convergence for covariance matrix estimation. *The Annals of Statistics* **38**(4).

Calvo, Eduard, Victor Martínez-de Albéniz. 2016. Sourcing strategies and supplier incentives for short-life-cycle goods. *Management Science* **62**(2) 436–455.

Carrasco, Vinicius, Vitor Farinha Luz, Nenad Kos, Matthias Messner, Paulo Monteiro, Humberto Moreira. 2018. Optimal selling mechanisms under moment conditions. *Journal of Economic Theory* **177** 245–279.

Catoni, Olivier. 2012. Challenging the empirical mean and empirical variance: A deviation study. *Annales de l'IHP Probabilités et statistiques* **48**(4) 1148–1185.

Cerreia-Vioglio, Simone, Lars Peter Hansen, Fabio Maccheroni, Massimo Marinacci. 2025. Making decisions under model misspecification. *Forthcoming in Review of Economic Studies.*

Chen, Hongqiao, Ming Hu, Georgia Perakis. 2022. Distribution-free pricing. *Manufacturing & Service Operations Management* **24**(4) 1939–1958.

Chen, Li, Melvyn Sim. 2025. Robust CARA optimization. *Operations Research* **73**(3) 1459–1478.

Chen, Xi, Akshay Krishnamurthy, Yining Wang. 2024a. Robust dynamic assortment optimization in the presence of outlier customers. *Operations Research* **72**(3) 999–1015.

Chen, Xin, Melvyn Sim, David Simchi-Levi, Peng Sun. 2007. Risk aversion in inventory management. *Operations Research* **55**(5) 828–842.

Chen, Zhi, Zhenyu Hu, Ruiqin Wang. 2024b. Screening with limited information: A dual perspective. *Operations Research* **72**(4) 1487–1504.

Chen, Zhi, Weijun Xie. 2021. Regret in the newsvendor model with demand and yield randomness. *Production and Operations Management* **30**(11) 4176–4197.

Chu, Leon Yang, Qi Feng, J George Shanthikumar, Zuo-Jun Max Shen, Jian Wu. 2025. Solving the price-setting newsvendor problem with parametric operational data analytics (ODA). *Management Science* **71**(8) 6627–6646.

Das, Bikramjit, Anulekha Dhara, Karthik Natarajan. 2021. On the heavy-tail behavior of the distributionally robust newsvendor. *Operations Research* **69**(4) 1077–1099.

De Meyer, Arnoud, Christoph Loch, Michael Pich. 2002. Managing project uncertainty: From variation to chaos. *MIT Sloan Management Review* **43**(2) 60–67.

Delage, Erick, Yinyu Ye. 2010. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research* **58**(3) 595–612.

Durrett, Rick. 2019. *Probability: Theory and examples*. Cambridge University Press.

Embrechts, Paul, Alexander Schied, Ruodu Wang. 2022. Robustness in the optimization of risk measures. *Operations Research* **70**(1) 95–110.

Feiler, Daniel, Jordan Tong. 2022. From noise to bias: Overconfidence in new product forecasting. *Management Science* **68**(6) 4685–4702.

Fu, Mingyang, Xiaobo Li, Lianmin Zhang. 2024. Distributionally robust newsvendor under stochastic dominance with a feature-based application. *Manufacturing & Service Operations Management* **26**(5) 1962–1977.

Gallego, Guillermo, Ilkyeong Moon. 1993. The distribution free newsboy problem: Review and extensions. *Journal of the Operational Research Society* **44**(8) 825–834.

Gelbrich, Matthias. 1990. On a formula for the L^2 Wasserstein metric between measures on Euclidean and Hilbert spaces. *Mathematische Nachrichten* **147**(1) 185–203.

Govindarajan, Aravind, Amitabh Sinha, Joline Uichanco. 2021. Distribution-free inventory risk pooling in a multilocation newsvendor. *Management Science* **67**(4) 2272–2291.

Han, Qiaoming, Donglei Du, Luis Zuluaga. 2014. A risk-and ambiguity-averse extension of the max-min newsvendor order formula. *Operations Research* **62**(3) 535–542.

Hanasusanto, Grani, Daniel Kuhn, Stein Wallace, Steve Zymler. 2015. Distributionally robust multi-item newsvendor problems with multimodal demand distributions. *Mathematical Programming* **152**(1-2) 1–32.

Hansen, Lars. 2014. Nobel lecture: Uncertainty outside and inside economic models. *Journal of Political Economy* **122**(5) 945–987.

Hansen, Lars, Thomas Sargent. 2022. Structured ambiguity and model misspecification. *Journal of Economic Theory* **199**(2) 105–165.

Keskin, Bora, Assaf Zeevi. 2017. Chasing demand: Learning and earning in a changing environment. *Mathematics of Operations Research* **42**(2) 277–307.

Li, Lujun, Hui Shao, Ruodu Wang, Jingping Yang. 2018. Worst-case range value-at-risk with partial information. *SIAM Journal on Financial Mathematics* **9**(1) 190–218.

Liu, Feng, Zhi Chen, Shuming Wang. 2023. Globalized distributionally robust counterpart. *INFORMS Journal on Computing* **35**(5) 1120–1142.

Liu, Peng, Alexander Schied, Ruodu Wang. 2021. Distributional transforms, probability distortions, and their applications. *Mathematics of Operations Research* **46**(4) 1490–1512.

Long, Daniel Zhuoyu, Melvyn Sim, Minglong Zhou. 2023. Robust satisficing. *Operations Research* **71**(1) 61–82.

Meyer, Jack. 1987. Two-moment decision models and expected utility maximization. *American Economic Review* **77**(3) 421–430.

Müller, Alfred, Marco Scarsini, Ilia Tsetlin, Robert Winkler. 2022. Ranking distributions when only means and variances are known. *Operations Research* **70**(5) 2851–2859.

Nambiar, Mila, David Simchi-Levi, He Wang. 2019. Dynamic learning and pricing with model misspecification. *Management Science* **65**(11) 4980–5000.

Natarajan, Karthik, Melvyn Sim, Chung-Piaw Teo. 2011. Beyond risk: Ambiguity in supply chains. *The Handbook of Integrated Risk Management in Global Supply Chains* 103–124.

Natarajan, Karthik, Melvyn Sim, Joline Uichanco. 2018. Asymmetry and ambiguity in newsvendor models. *Management Science* **64**(7) 3146–3167.

Nguyen, Viet, Soroosh Shafieezadeh-Abadeh, Damir Filipović, Daniel Kuhn. 2021. Mean-covariance robust risk measurement. *Available at* https://arxiv.org/abs/2112.09959.

Olivares, Marcelo, Christian Terwiesch, Lydia Cassorla. 2008. Structural estimation of the newsvendor model: An application to reserving operating room time. *Management Science* **54**(1) 41–55.

Padmanabhan, Divya, Karthik Natarajan, Karthyek Murthy. 2021. Exploiting partial correlations in distributionally robust optimization. *Mathematical Programming* **186**(1) 209–255.

Perakis, Georgia, Guillaume Roels. 2008. Regret in the newsvendor model with partial information. *Operations Research* **56**(1) 188–203.

Petracou, Electra, Anastasios Xepapadeas, Athanasios Yannacopoulos. 2022. Decision making under model uncertainty: Fréchet-Wasserstein mean preferences. *Management Science* **68**(2) 1195–1211.

Popescu, Ioana. 2007. Robust mean-covariance solutions for stochastic optimization. *Operations Research* **55**(1) 98–112.

Qi, Meng, Zuo-Jun Max Shen. 2022. Integrating prediction/estimation and optimization with applications in operations management. *Tutorials in Operations Research: Emerging and Impactful Topics in Operations*. INFORMS, 36–58.

Quiñonero-Candela, Joaquin, Sugiyama Masashi, Schwaighofer Anton, Neil Lawrence. 2008. *Dataset shift in machine learning*. MIT Press.

Rahimian, Hamed, Güzin Bayraksan, Tito Homem-de Mello. 2019. Controlling risk and demand ambiguity in newsvendor models. *European Journal of Operational Research* **279**(3) 854–868.

Scarf, Herbert. 1958. A min-max solution of an inventory problem. *Studies in the mathematical theory of inventory and production.*

Simchi-Levi, David, Yehua Wei. 2015. Worst-case analysis of process flexibility designs. *Operations Research* **63**(1) 166–185.

Sugiyama, Masashi, Motoaki Kawanabe. 2012. *Machine learning in non-stationary environments: Introduction to covariate shift adaptation*. MIT press.

Sutter, Tobias, Andreas Krause, Daniel Kuhn. 2021. Robust generalization despite distribution shift via minimum discriminating information. *Advances in Neural Information Processing*

Systems **34** 29754–29767.

Urban, Glen, John Hauser. 1993. *Design and Marketing of New Products*. 2nd ed. Pearson.

Vershynin, Roman. 2010. Introduction to the non-asymptotic analysis of random matrices. *Available at* https://arxiv.org/abs/1011.3027.

Villani, Cédric. 2009. *Optimal transport: Old and new*. Springer.

Wiesemann, Wolfram, Daniel Kuhn, Melvyn Sim. 2014. Distributionally robust convex optimization. *Operations Research* **62**(6) 1358–1376.

Wozabal, David. 2014. Robustifying convex risk measures for linear portfolios: A nonparametric approach. *Operations Research* **62**(6) 1302–1315.

Zhang, Luhao, Jincheng Yang, Rui Gao. 2023. Optimal robust policy for feature-based newsvendor. *Management Science* **70**(4) 2315–2329.

Zhu, Zhisu, Jiawei Zhang, Yinyu Ye. 2013. Newsvendor optimization with limited distribution information. *Optimization Methods and Software* **28**(3) 640–667.

E-Companion to "Newsvendor under Ambiguity and Misspecification"

EC.1. Technical Lemmas

LEMMA EC.1. *For any $F, G \in \mathcal{P}$, $d(F,G)$ in (2) is jointly convex in F and G.*

Proof. Given $F_1, F_2, G_1, G_2 \in \mathcal{P}$, let $\Gamma_1 \in \mathcal{W}(F_1, G_1)$ and $\Gamma_2 \in \mathcal{W}(F_2, G_2)$ be the joint distributions that solve the corresponding optimal transport. That is, $d(F_1,G_1) = \int_{\mathbb{R}_+\times\mathbb{R}_+} |u-v|^2 \Gamma_1(\mathrm{d}u,\mathrm{d}v)$ and $d(F_2,G_2) = \int_{\mathbb{R}_+\times\mathbb{R}_+} |u-v|^2 \Gamma_2(\mathrm{d}u,\mathrm{d}v)$. We first show that the joint distribution $\Gamma_\lambda = (1-\lambda)\Gamma_1 + \lambda\Gamma_2$ has marginals $F_\lambda = (1-\lambda)F_1 + \lambda F_2$ and $G_\lambda = (1-\lambda)G_1 + \lambda G_2$. In fact, for any Borel set $\mathfrak{B} \subseteq \mathbb{R}_+$, we have

$$\Gamma_\lambda(\mathfrak{B}\times\mathbb{R}_+) = (1-\lambda)\Gamma_1(\mathfrak{B}\times\mathbb{R}_+) + \lambda\Gamma_2(\mathfrak{B}\times\mathbb{R}_+) = (1-\lambda)F_1(\mathfrak{B}) + \lambda F_2(\mathfrak{B})$$
$$\Gamma_\lambda(\mathbb{R}_+\times\mathfrak{B}) = (1-\lambda)\Gamma_1(\mathbb{R}_+\times\mathfrak{B}) + \lambda\Gamma_2(\mathbb{R}_+\times\mathfrak{B}) = (1-\lambda)G_1(\mathfrak{B}) + \lambda G_2(\mathfrak{B}),$$

which implies that $\Gamma_\lambda \in \mathcal{W}(F_\lambda, G_\lambda)$. As a result,

$$d(F_\lambda, G_\lambda) \le \int_{\mathbb{R}_+\times\mathbb{R}_+} |u-v|^2\, \Gamma_\lambda(\mathrm{d}u,\mathrm{d}v) = (1-\lambda)\cdot d(F_1,G_1) + \lambda\cdot d(F_2,G_2),$$

which concludes the proof. □

LEMMA EC.2. *Suppose that $0<\varepsilon<\infty$, and let $\Upsilon_\varepsilon^\star$ denote the optimal value of problem (3) corresponding to ε, and $\Pi_\alpha^\star$ the optimal value of* MISSPECIFICATION *corresponding to α. Then,*

$$\Upsilon_\varepsilon^\star = \max_{\alpha\ge 0}\{\Pi_\alpha^\star - \varepsilon\alpha\}. \tag{EC.1}$$

The optimal value $\alpha^\star$ of problem (EC.1) is achieved, and the optimal solution $(q_{\alpha^\star}^\star, F_{\alpha^\star}^\star)$ of MISSPECIFICATION *associated with $\alpha = \alpha^\star$ is also the optimal solution to problem (3).*

Proof. We start by fixing the order quantity. For any fixed $q \ge 0$, we define a function $\varphi(q,\varepsilon) = \min_{d(F,\mathcal{A})\le\varepsilon} \mathbb{E}_F[\pi(q,\tilde{v})]$. First, it is clear that $\varphi(q,\varepsilon)$ is decreasing in ε. Second, $\varphi(q,\varepsilon)$ is bounded on $\mathbb{R}_+$ because for any $\varepsilon\in(0,+\infty)$, $-cq \le \varphi(q,\varepsilon) \le (p-c)q$. Third, $\varphi(q,\varepsilon)$ is convex in ε on $\mathbb{R}_{++}$. To see this, we fix $F_1, F_2 \in \mathcal{P}$ and $\varepsilon_1, \varepsilon_2 \ge 0$ such that $d(F_1,\mathcal{A}) \le \varepsilon_1$ and $d(F_2,\mathcal{A}) \le \varepsilon_2$. Note that $d(F,\mathcal{A}) = \min_{G\in\mathcal{A}} d(F,G)$ is convex in F because $d(F,G)$ is jointly convex in F and G (Lemma EC.1) and maximization over a convex set preserves convexity. For any $\lambda\in[0,1]$, it holds that $d(\lambda F_1 + (1-\lambda)F_2, \mathcal{A}) \le \lambda d(F_1,\mathcal{A}) + (1-\lambda)d(F_2,\mathcal{A}) \le \lambda\varepsilon_1 + (1-\lambda)\varepsilon_2$, where the first inequality follows from the convexity of $d(\cdot,\mathcal{A})$. Thus we have $\varphi(q, \lambda\varepsilon_1 + (1-\lambda)\varepsilon_2) \ge \mathbb{E}_{\lambda F_1 + (1-\lambda)F_2}[\pi(q,\tilde{u})] = \lambda\cdot\mathbb{E}_{F_1}[\pi(q,\tilde{u})] + (1-\lambda)\cdot\mathbb{E}_{F_2}[\pi(q,\tilde{u})]$. Taking the minimum over F_1 and F_2 yields $\varphi(q,\lambda\varepsilon_1 + (1-\lambda)\varepsilon_2) \ge \lambda\varphi(q,\varepsilon_1) + (1-\lambda)\varphi(q,\varepsilon_2)$, which establishes the convexity (and continuity) of $\varphi(q,\varepsilon)$. Finally, given $\alpha>0$, the Legendre transform of the convex function $\varphi(q,\cdot)$ is

$$\varphi^*(q,\alpha) = \min_{\varepsilon\ge 0}\{\alpha\varepsilon + \varphi(q,\varepsilon)\} = \min_{\varepsilon\ge 0}\min_{F\in\mathcal{P}}\{\alpha\varepsilon + \mathbb{E}_F[\pi(q,\tilde{u})] : d(F,\mathcal{A})\le\varepsilon\} = \min_{F\in\mathcal{P}}\{\alpha\cdot d(F,\mathcal{A}) + \mathbb{E}_F[\pi(q,\tilde{u})]\},$$

which is concave in α. Note that the above relation also holds for $\alpha = 0$. For any $\varepsilon>0$, applying Legendre transform on the concave function $\varphi^*(q,\cdot)$ yields

$$(\varphi^*(q,\varepsilon))^* = \max_{\alpha\ge 0}\{\varphi^*(q,\alpha) - \alpha\varepsilon\} = \max_{\alpha\ge 0}\min_{F\in\mathcal{P}}\{\mathbb{E}_F[\pi(q,\tilde{u})] + \alpha\cdot d(F,\mathcal{A}) - \varepsilon\alpha\}.$$

Since $\varphi(q,\varepsilon)$ is bounded, convex, and continuous, for any $\varepsilon>0$, it holds that $\varphi(q,\varepsilon) = (\varphi^*(q,\varepsilon))^*$. We now optimize the order quantity. Maximizing over $q\ge 0$ yields

$$\Upsilon_\varepsilon^\star = \max_{q\ge 0}\varphi(q,\varepsilon) = \max_{q\ge 0}\min_{d(F,\mathcal{A})\le\varepsilon}\mathbb{E}_F[\pi(q,\tilde{u})] \tag{EC.2}$$
$$= \max_{\alpha\ge 0}\max_{q\ge 0}\Big\{\min_{F\in\mathcal{P}}\{\mathbb{E}_F[\pi(q,\tilde{u})] + \alpha\cdot d(F,\mathcal{A})\} - \varepsilon\alpha\Big\} \tag{EC.3}$$
$$= \max_{\alpha\ge 0}\{\Pi_\alpha^\star - \varepsilon\alpha\}, \tag{EC.4}$$

where (EC.4) follows from the definition of $\Pi_\alpha^\star$. Indeed, the optimal value of (EC.4) is achieved at some finite $\alpha^\star \ge 0$ because the function $\Pi_\alpha^\star - \varepsilon\alpha$ is continuous in α with $\Pi_0^\star - \varepsilon\cdot 0 = 0$ and $\Pi_\alpha^\star - \varepsilon\alpha \to -\infty$ as $\alpha\to\infty$. Finally, note that the optimal solution $(\alpha^\star, q^\star, F^\star)$ of (EC.3) has its part $(q^\star, F^\star)$ being the optimal solution of MISSPECIFICATION with $\alpha = \alpha^\star$, which we have denoted by $(q_\alpha^\star, F_\alpha^\star)$. Hence, $(q_{\alpha^\star}^\star, F_{\alpha^\star}^\star)$ is optimal to (EC.2), *i.e.*, problem (3). This concludes the proof. □

Lemma EC.3 (Interchangeability, Zhang et al. 2025). *Given $G \in \mathcal{P}_M$ and a G-measurable function $h(\cdot): \mathbb{R}_+^M \mapsto \mathbb{R}$ with $\mathbb{E}_G[h(\tilde{\boldsymbol{u}})] < +\infty$, we have*

$$\min_{F \in \mathcal{P}_M, \Gamma \in \mathcal{W}(F,G)} \mathbb{E}_\Gamma\left[h(\tilde{\boldsymbol{u}}) + \alpha \cdot \|\tilde{\boldsymbol{u}} - \tilde{\boldsymbol{v}}\|_2^2\right] = \mathbb{E}_G\left[\min_{\boldsymbol{u} \geq 0} \left\{h(\boldsymbol{u}) + \alpha \cdot \|\boldsymbol{u} - \tilde{\boldsymbol{v}}\|_2^2\right\}\right].$$

Proof. Note that the $(\mathbb{R}_+^M, \|\cdot\|_2)$ is a metric space and every Borel probability measure $G \in \mathcal{P}(\mathbb{R}_+^M)$ is tight. The result is then an immediate consequence of proposition 2 in Zhang et al. (2025). □

Lemma EC.4. *Let $\ell(\alpha, q, v) = \min_{u \geq 0} \{\pi(q,u) + \alpha(u-v)^2\}$. For any fixed $v \geq 0$,*

(i) if $0 \leq q \leq \frac{p}{4\alpha}$, then $\ell(\alpha,q,v) = \min\{\alpha v^2, pq\} - cq$;

(ii) if $q > \frac{p}{4\alpha}$, then

$$\ell(\alpha,q,v) = \begin{cases} \alpha v^2 - cq & 0 \leq v \leq \frac{p}{2\alpha} \\ p \cdot \min\left\{v - \frac{p}{4\alpha}, q\right\} - cq & v > \frac{p}{2\alpha}. \end{cases}$$

Proof. The function $\pi(q,v) + \alpha(v-u)^2$ can be written as a piecewise quadratic function as follows:

$$g(u,q,v) = p\min\{q,u\} - cq + \alpha(u-v)^2 = \begin{cases} \underline{g}(u) = pu - cq + \alpha(u-v)^2 & u \leq q \\ \bar{g}(u) = pq - cq + \alpha(u-v)^2 & u > q. \end{cases}$$

Before proceeding, we denote the left and right derivatives of $g(u,q,v)$ at $u = u_0$ by $g'_-(u_0)$ and $g'_+(u_0)$, respectively. Based on the value of v, there are three cases to consider.

(*i*) If $v \geq q + \frac{p}{2\alpha}$, *i.e.*, $\underline{g}'_-(q) = 2\alpha(q-v) + p \leq 0$, then $\underline{g}(u)$ is decreasing over $(0,q]$. Note that $\bar{g}'(u) = 2\alpha(u-v) = 0$ admits a unique solution $u = v$, implying that $\bar{g}(u)$ is decreasing in (q,v) and increasing in $[v,+\infty)$. Therefore, it holds that $\min_{u\geq 0} g(u,q,v) = \bar{g}(v) = pq - cq$.

(*ii*) If $v \leq q$, *i.e.*, $\bar{g}'_+(q) = 2\alpha(q-v) \geq 0$, and we have $\underline{g}'(u) = 2\alpha(u-v) + p = 0$ admits a unique solution $u = v - \frac{p}{2\alpha}$, implying that $\underline{g}(u)$ is first decreasing in $[0,(v-p/2\alpha)^+]$ and then increasing in $((v-p/2\alpha)^+, q]$, and $\bar{g}(u)$ is increasing in $(q,+\infty)$. Therefore, it holds that $\min_{u\geq 0} g(u,q,v) = \underline{g}((v-\frac{p}{2\alpha})^+)$.

(*iii*) If $q \leq v \leq q + \frac{p}{2\alpha}$, *i.e.*, $\underline{g}'_-(q) \geq 0$ and $\bar{g}'_+(q) \leq 0$, then both $\underline{g}(u)$ and $\bar{g}(u)$ admit a corresponding minimizer in the domain. It follows that $\min_{u\geq 0} g(u,q,v) = \min\{\underline{g}((v-\frac{p}{2\alpha})^+), \bar{g}(v)\}$.

Note that when $v < \frac{p}{2\alpha}$, we have $\underline{g}\big((v-\frac{p}{2\alpha})^+\big) = \alpha v^2 - cq$ and when $\frac{p}{2\alpha} \leq v \leq q + \frac{p}{2\alpha}$, $\underline{g}\big((v-\frac{p}{2\alpha})^+\big) = p(v - \frac{p}{4\alpha}) - cq$. We next discuss based on the value of q. When $q \geq \frac{p}{2\alpha}$, we have

$$\ell(\alpha,q,v) = \min_{u\geq 0}\ g(u,q,v) = \begin{cases} \alpha v^2 - cq & v \leq \frac{p}{2\alpha} \\ p(v - \frac{p}{4\alpha}) - cq & \frac{p}{2\alpha} \leq v \leq q \\ p \cdot \min\{v - \frac{p}{4\alpha}, q\} - cq & q \leq v \leq q + \frac{p}{2\alpha} \\ pq - cq & v \geq q + \frac{p}{2\alpha} \end{cases} = \begin{cases} \alpha v^2 - cq & v \leq \frac{p}{2\alpha} \\ p(v - \frac{p}{4\alpha}) - cq & \frac{p}{2\alpha} \leq v \leq q + \frac{p}{4\alpha} \\ pq - cq & v \geq q + \frac{p}{4\alpha}; \end{cases}$$

When $q \leq \frac{p}{2\alpha}$, we have

$$\ell(\alpha,q,v) = \min_{u\geq 0}\ g(u,q,v) = \begin{cases} \alpha v^2 - cq & v \leq q \\ \min\{\alpha v^2, pq\} - cq & q \leq v \leq \frac{p}{2\alpha} \\ p \cdot \min\{v - \frac{p}{4\alpha}, q\} - cq & \frac{p}{2\alpha} \leq v \leq q + \frac{p}{2\alpha} \\ pq - cq & v \geq q + \frac{p}{2\alpha}. \end{cases}$$

If $\frac{p}{4\alpha} \leq q \leq \frac{p}{2\alpha}$, then $\min\{\alpha v^2, pq\} = \alpha v^2$ for $u \in \left[q, \frac{p}{2\alpha}\right]$, resulting in

$$\ell(\alpha,q,v) = \begin{cases} \alpha v^2 - cq & v \leq \frac{p}{2\alpha} \\ p(v - \frac{p}{4\alpha}) - cq & \frac{p}{2\alpha} \leq v \leq q + \frac{p}{4\alpha} \\ pq - cq & v \geq q + \frac{p}{4\alpha}. \end{cases}$$

If $q \leq \frac{p}{4\alpha}$, then $p \cdot \min\left\{v - \frac{p}{4\alpha}, q\right\} - cq = pq - cq$ for $v \in \left[\frac{p}{2\alpha}, q + \frac{p}{2\alpha}\right]$. Correspondingly,

$$\ell(\alpha,q,v) = \begin{cases} \alpha v^2 - cq & v \leq \sqrt{\frac{pq}{\alpha}} \\ pq - cq & v \geq \sqrt{\frac{pq}{\alpha}} \end{cases} = \min\{\alpha v^2, pq\} - cq.$$

Consolidating these results based on the three ranges of q then completes the proof. □

EC.2. Proofs.

Proof of Theorem 1. In the MISSPECIFICATION problem, given $q \geq 0$ and $G \in \mathcal{A}$, it holds that

$$\min_{F\in\mathcal{P}} \{\mathbb{E}_F[\pi(q,\tilde{u})] + \alpha \cdot d(F,G)\} = \mathbb{E}_G\big[\min_{u\geq 0}\{\pi(q,u) + \alpha(u-\tilde{v})^2\}\big] = \mathbb{E}_G[\ell(\alpha,q,\tilde{v})].$$

Here, the first equality follows from the interchangeability principle (Lemma EC.3) and the second equality follows from $\ell(\alpha,q,v) = \min_{u\geq 0}\{\pi(q,u) + \alpha(u-v)^2\}$ (see Lemma EC.4 for its closed-form expression). It suffices to verify for any $q \geq 0$ that $\int_{\mathbb{R}_+} \pi(q,v)\mathrm{d}T_{\varphi_\alpha}[G](v) = \int_{\mathbb{R}_+} \ell(\alpha,q,v)\mathrm{d}G(v)\ \ \forall G \in \mathcal{A}$. In view of (i), that is, $q < \frac{p}{4\alpha}$, we have $\int_{\mathbb{R}_+} \pi(q,v)\mathrm{d}T_{\varphi_\alpha}[G](v) = \int_{\mathbb{R}_+} \pi(q,v)\mathrm{d}G\big(\sqrt{\tfrac{p}{\alpha}v}\big) = \int_{\mathbb{R}_+} \pi\big(q,\tfrac{\alpha}{p}v^2\big)\mathrm{d}G(v) = \int_{\mathbb{R}_+} \ell(\alpha,q,v)\mathrm{d}G(v)$, where the second equality follows from the variable substitution $v \leftarrow \sqrt{\frac{p}{\alpha}v}$ and the third equality follows from the fact that $\ell(\alpha,q,v) = \min\{\alpha v^2, pq\} - cq$ when $q < \frac{p}{4\alpha}$.

As for (ii), that is, $q \geq \frac{p}{4\alpha}$, we have

$$\begin{aligned}\int_{\mathbb{R}_+} \pi(q,v)\,\mathrm{d}T_{\varphi_\alpha}[G](v) &= \int_0^{\frac{p}{4\alpha}} \pi(q,v)\mathrm{d}G\big(\sqrt{\tfrac{p}{\alpha}v}\big) + \int_{\frac{p}{4\alpha}}^{\infty} \pi(q,v)\mathrm{d}G\big(v+\tfrac{p}{4\alpha}\big)\\ &= \int_0^{\frac{p}{2\alpha}} \pi\big(q,\tfrac{\alpha}{p}v^2\big)\mathrm{d}G(v) + \int_{\frac{p}{2\alpha}}^{\infty} \pi\big(q,v-\tfrac{p}{4\alpha}\big)\mathrm{d}G(v)\\ &= \int_0^{\infty} \ell(\alpha,q,v)\mathrm{d}G(v),\end{aligned}$$

where the second equality follows from the variable substitution: $v \leftarrow \sqrt{\frac{p}{\alpha}v}$ for $v < \frac{p}{4\alpha}$ and $v \leftarrow v + \frac{p}{4\alpha}$ for $v \geq \frac{p}{4\alpha}$ and the third equality follows from the fact that $\ell(\alpha,q,v) = \min\{\alpha v^2, pq\} - cq = \alpha v^2 - cq$ when $v < \frac{p}{2\alpha}$ and $\ell(\alpha,q,v) = p\min\big\{v - \frac{p}{4\alpha}, q\big\} - cq$ when $v \geq \frac{p}{2\alpha}$. □

Proof of Proposition 1. Note that for any $q \geq 0$, it holds that $L(q) = \min_{F\in\mathcal{P},\,G\in\mathcal{A}} \big\{\mathbb{E}_F[\pi(q,\tilde{u})] + \alpha \cdot d(F,G)\big\} = \min_{G\in\mathcal{A}} \mathbb{E}_G[\Psi(\alpha,q,\tilde{v})]$, where $\Psi(\alpha,q,v) = \pi(q,\varphi_\alpha(v))$. Given $q \geq 0$, $L(q)$ is a moment problem:

$$\begin{aligned}\min_{G}\ & \int_{\mathbb{R}_+} \Psi(\alpha,q,v)\,\mathrm{d}G(v)\\ \text{s.t.}\ & \int_{\mathbb{R}_+} v\,\mathrm{d}G(v) = \mu && \cdots\ s_\alpha\\ & \int_{\mathbb{R}_+} v^2\,\mathrm{d}G(v) = \mu^2+\sigma^2 && \cdots\ r_\alpha\\ & \int_{\mathbb{R}_+} \mathrm{d}G(v) = 1 && \cdots\ t_\alpha\\ & G \in \mathcal{M}_+,\end{aligned} \tag{PRIMAL}$$

whose dual is

$$\begin{aligned}\max_{s_\alpha,\,r_\alpha,\,t_\alpha}\ & \mu s_\alpha - (\mu^2+\sigma^2)r_\alpha - t_\alpha\\ \text{s.t.}\ & v s_\alpha - v^2 r_\alpha - t_\alpha \leq \Psi(\alpha,q,v)\quad \forall v \geq 0.\end{aligned} \tag{DUAL}$$

We next derive the expression of $L(q)$ by constructing a pair of primal and dual feasible solutions that attain the same objective value (*i.e.*, strong duality holds between PRIMAL and DUAL). We first find a dual feasible solution $(s_\alpha, r_\alpha, t_\alpha)$ such that the quadratic function $g(v) = vs_\alpha - v^2 r_\alpha - t_\alpha$ touches $\Psi(\alpha,q,v)$ at exactly two points, where there are three scenarios in total (as shown in Figure EC.1). We then find a primal feasible solution $G_\alpha \in \mathcal{A}$ whose objective value matches the dual objective under $(s_\alpha, r_\alpha, t_\alpha)$. This is achieved by setting the support of G_α as the tangent points of $g(v)$ and $\Psi(\alpha,q,v)$ and then solving for the corresponding probabilities based on the moment conditions in $\mathcal{A}$. In the following, we consider these three scenarios based on the value of q.

Scenario 1. When $0 \leq q \leq \frac{p}{4\alpha}$, we first construct a feasible distribution to PRIMAL as follows:

$$G_\alpha = \left(\frac{1}{2} - \frac{\mu^2-\sigma^2-\frac{pq}{\alpha}}{2\sqrt{(\frac{pq}{\alpha}+\mu^2+\sigma^2)^2 - 4\mu^2\frac{pq}{\alpha}}}\right)\cdot\delta_{v_1} + \left(\frac{1}{2} + \frac{\mu^2-\sigma^2-\frac{pq}{\alpha}}{2\sqrt{(\frac{pq}{\alpha}+\mu^2+\sigma^2)^2 - 4\mu^2\frac{pq}{\alpha}}}\right)\cdot\delta_{v_2}, \tag{EC.5}$$

where the support points are

$$v_1 = \frac{1}{2\mu}\Big(\frac{pq}{\alpha} + \mu^2 + \sigma^2 - \sqrt{(\tfrac{pq}{\alpha} + \mu^2+\sigma^2)^2 - 4\mu^2\tfrac{pq}{\alpha}}\Big) \text{ and } v_2 = \frac{1}{2\mu}\Big(\frac{pq}{\alpha} + \mu^2 + \sigma^2 + \sqrt{(\tfrac{pq}{\alpha} + \mu^2+\sigma^2)^2 - 4\mu^2\tfrac{pq}{\alpha}}\Big).$$

One can verify that $G_\alpha \in \mathcal{A}$ and the corresponding primal objective value under G_α is equal to $\mathbb{E}_{G_\alpha}[\Psi(\alpha,q,\tilde{u})] = \frac{\alpha}{2}\Big(\frac{pq}{\alpha} + \mu^2 + \sigma^2 - \sqrt{(\frac{pq}{\alpha} + \mu^2+\sigma^2)^2 - 4\mu^2\frac{pq}{\alpha}}\Big) - cq$. We next construct a dual feasible solution that attains the same dual objective value. Note that $\Psi(\alpha,q,v) = \min\{\alpha v^2, pq\} - cq\ \forall v \geq 0$ when $0 \leq q \leq \frac{p}{4\alpha}$ (Lemma EC.4). Hence, DUAL becomes

$$\begin{aligned}\max_{s_\alpha,\,r_\alpha,\,t_\alpha}\ & s_\alpha\mu - r_\alpha(\mu^2+\sigma^2) - t_\alpha\\ \text{s.t.}\ & s_\alpha v - r_\alpha v^2 - t_\alpha \leq \alpha v^2 - cq && \forall v \geq 0\\ & s_\alpha v - r_\alpha v^2 - t_\alpha \leq pq - cq && \forall v \geq 0.\end{aligned}$$

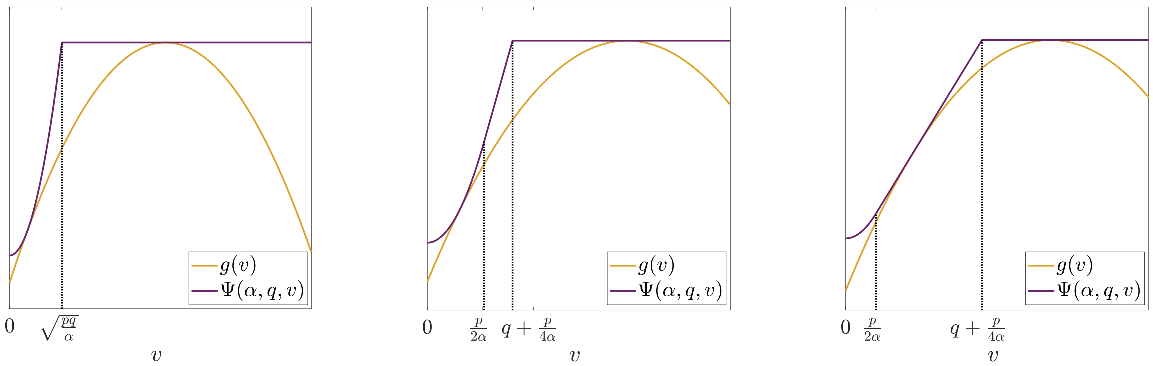


Figure EC.1 Illustration of the dual problem. Here, $g(v) = vs_\alpha - v^2 r_\alpha - t_\alpha$.

Consider the following solution:

$$s_\alpha = \frac{2\mu pq}{\sqrt{(\frac{pq}{\alpha}+\mu^2+\sigma^2)^2 - 4\mu^2\frac{pq}{\alpha}}}, r_\alpha = \frac{\alpha}{2}\left(\frac{\frac{pq}{\alpha}+\mu^2+\sigma^2}{\sqrt{(\frac{pq}{\alpha}+\mu^2+\sigma^2)^2 - 4\mu^2\frac{pq}{\alpha}}} - 1\right), t_\alpha = \frac{pq}{2}\left(\frac{\frac{pq}{\alpha}+\mu^2+\sigma^2}{\sqrt{(\frac{pq}{\alpha}+\mu^2+\sigma^2)^2 - 4\mu^2\frac{pq}{\alpha}}} - 1\right) + cq, \quad \text{(EC.6)}$$

which satisfies $s_\alpha \mu - r_\alpha(\mu^2+\sigma^2) - t_\alpha = \frac{\alpha}{2}\big(\frac{pq}{\alpha} + \mu^2 + \sigma^2 - \sqrt{(\frac{pq}{\alpha}+\mu^2+\sigma^2)^2 - 4\mu^2\frac{pq}{\alpha}}\big) - cq$. It remains to argue that this solution is feasible to DUAL. If $q = 0$, then $s_\alpha = r_\alpha = t_\alpha = 0$, naturally feasible to DUAL. We next investigate $q > 0$. In this case, $r_\alpha > 0$. The first semi-infinite constraint of DUAL is equivalent to $\max_{v\geq 0}\{s_\alpha v - (r_\alpha+\alpha)v^2 - t_\alpha + cq\} \leq 0$. For the left-hand side maximization, the optimal solution is $v^\star = \frac{s_\alpha}{2(r_\alpha+\alpha)} \geq 0$, which attains an optimal value of $\frac{s_\alpha^2}{4(r_\alpha+\alpha)} - t_\alpha + cq = 0$. Hence, the first semi-infinite constraint is satisfied. Similarly, the second semi-infinite constraint of DUAL is equivalent to $\max_{v\geq 0}\{s_\alpha v - r_\alpha v^2 - t_\alpha - pq + cq\} \leq 0$. For the left-hand side, the optimal solution is $v^\star = \frac{s_\alpha}{2r_\alpha}$ and the corresponding optimal value is $\frac{s_\alpha^2}{4r_\alpha} - t_\alpha - pq + cq = \frac{pq(r_\alpha+\alpha)}{\alpha} - \frac{pq}{\alpha}r_\alpha - pq = 0$. Hence, the second semi-infinite constraint is also satisfied, concluding that solution (EC.6) is feasible to DUAL and establishing the strong duality.

Scenario 2. When $q \geq \frac{p}{4\alpha}$ and $(2\mu - \frac{p}{\alpha})q < \mu^2 + \sigma^2 - \frac{p\mu}{2\alpha}$, DUAL becomes

$$\begin{aligned}
\max_{s_\alpha, r_\alpha, t_\alpha} \quad & s_\alpha\mu - r_\alpha(\mu^2+\sigma^2) - t_\alpha \\
\text{s.t.} \quad & s_\alpha v - r_\alpha v^2 - t_\alpha \leq \alpha v^2 - cq && \forall 0 \leq v \leq \tfrac{p}{2\alpha} \\
& s_\alpha v - r_\alpha v^2 - t_\alpha \leq p\big(v - \tfrac{p}{4\alpha}\big) - cq && \forall v \geq \tfrac{p}{2\alpha} \\
& s_\alpha v - r_\alpha v^2 - t_\alpha \leq pq - cq && \forall v \geq \tfrac{p}{2\alpha}.
\end{aligned}$$

Consider the pair of primal feasible solution (EC.5) and dual solution (EC.6). Upon the results in *Step 1*, we can establish the feasibility of the constructed dual solution similar to *Scenario 1*.

Scenario 3. When $q \geq \frac{p}{4\alpha}$ and $(2\mu - \frac{p}{\alpha})q \geq \mu^2 + \sigma^2 - \frac{p\mu}{2\alpha}$, we first construct a feasible primal solution:

$$G_\alpha = \frac{1}{2}\left(1 + \frac{q + \frac{p}{4\alpha} - \mu}{\sqrt{(q+\frac{p}{4\alpha}-\mu)^2+\sigma^2}}\right)\cdot\delta_{v_1} + \frac{1}{2}\left(1 - \frac{q + \frac{p}{4\alpha} - \mu}{\sqrt{(q+\frac{p}{4\alpha}-\mu)^2+\sigma^2}}\right)\cdot\delta_{v_2}, \quad \text{(EC.7)}$$

where $v_1 = q + \frac{p}{4\alpha} - \sqrt{(q+\frac{p}{4\alpha}-\mu)^2+\sigma^2}$ and $v_2 = q + \frac{p}{4\alpha} + \sqrt{(q+\frac{p}{4\alpha}-\mu)^2+\sigma^2}$. One can verify that $G_\alpha \in \mathcal{A}$ and $\mathbb{E}_{G_\alpha}[\Psi(\alpha, q, \tilde{u})] = \frac{p}{2}\big(\mu - q - \frac{p}{4\alpha} - \sqrt{(q+\frac{p}{4\alpha}-\mu)^2+\sigma^2}\big) + (p-c)q$. We next construct a dual feasible solution that attains the same dual objective value. Consider the following solution:

$$s_\alpha = \frac{p}{2} + 2r_\alpha\big(q + \frac{p}{4\alpha}\big), t_\alpha = \frac{p^2}{16r_\alpha} + r_\alpha\big(q+\frac{p}{4\alpha}\big)^2 + \frac{p}{2}\big(q + \frac{p}{4\alpha}\big) - (p-c)q, r_\alpha = \frac{p}{4\sqrt{(q+\frac{p}{4\alpha}-\mu)^2+\sigma^2}}, \quad \text{(EC.8)}$$

which is feasible to DUAL and attains the same primal objective value under G_α.

Note that *Scenario 1* and *Scenario 2* correspond to the same objective function. Combining the results in these three scenarios, we then obtain the desired result immediately. □

Proof of Theorem 2. Defining $L(q) = \min_{G\in\mathcal{A}} \mathbb{E}_G[\Psi(\alpha, q, \tilde{v})]$, MISSPECIFICATION is then equivalent to $\max_{q\geq 0} L(q)$. Note that when $\kappa < \frac{\sigma^2}{\mu^2+\sigma^2}$, $\max_{q\geq 0}\min_{G\in\mathcal{A}}\mathbb{E}_G[\Psi(\alpha,q,\tilde{v})] \leq \max_{q\geq 0}\min_{G\in\mathcal{A}}\mathbb{E}_G[\pi(q,\tilde{v})] = 0$, where the inequality holds since $\Psi(\alpha, q, v) = \pi(q, \varphi_\alpha(v)) \leq \pi(q, v)\ \forall q \geq 0$ and the equality follows from the tight lower bound derived in Scarf (1958). The order quantity $q = 0$ satisfies $\mathbb{E}_G[\Psi(\alpha, 0, \tilde{v})] = 0\ \forall G \in \mathcal{A}$ (*i.e.*,

$\min_{G\in\mathcal{A}} \mathbb{E}_G[\Psi(\alpha,0,\tilde{v})]=0$). Hence, $q^\star_\alpha=0$. This corresponds to case (*ii*) in the statement. In the remainder of the proof, we focus on $\kappa\geq\frac{\sigma^2}{\mu^2+\sigma^2}$. As shown in Proposition 1, we have

$$L(q)=\begin{cases}\frac{p}{2}\big(\mu-q-\frac{p}{4\alpha}-\sqrt{(q+\frac{p}{4\alpha}-\mu)^2+\sigma^2}\big)+(p-c)q & \text{if } q\in\mathcal{Q}\\ \frac{\alpha}{2}\big(\frac{pq}{\alpha}+\mu^2+\sigma^2-\sqrt{(\frac{pq}{\alpha}+\mu^2+\sigma^2)^2-4\mu^2\frac{pq}{\alpha}}\big)-cq & \text{otherwise}\end{cases}$$

where $\mathcal{Q}=\{q\in\mathbb{R}_+ \mid q\geq\frac{p}{4\alpha},\ (2\mu-\frac{p}{\alpha})q\geq\mu^2+\sigma^2-\frac{p\mu}{2\alpha}\}$. Based on the value of α, we divide the arguments into three different scenarios.

Scenario 1. Suppose that $\alpha<\frac{p}{2\mu}$, *i.e.*, $2\mu-\frac{p}{\alpha}<0$. For any $q\geq\frac{p}{4\alpha}$, it holds that $(2\mu-\frac{p}{\alpha})q\leq(2\mu-\frac{p}{\alpha})\frac{p}{4\alpha}=\frac{p}{\alpha}(\mu-\frac{p}{4\alpha})-\frac{p\mu}{2\alpha}<\mu^2+\sigma^2-\frac{p\mu}{2\alpha}$, where the strict inequality follows from the fact that $\frac{p}{\alpha}(\mu-\frac{p}{4\alpha})<\mu^2$ when $\alpha<\frac{p}{2\mu}$. That is to say, for any $q\geq0$ we have $L(q)=\frac{\alpha}{2}\big(\frac{pq}{\alpha}+\mu^2+\sigma^2-\sqrt{(\frac{pq}{\alpha}+\mu^2+\sigma^2)^2-4\mu^2\frac{pq}{\alpha}}\big)-cq$. Setting the derivative of $L(q)$ to 0, we then obtain $q^\star_\alpha=\big(\mu^2-\sigma^2+2f(1-\kappa)\mu\sigma\big)\cdot\frac{\alpha}{p}$.

Scenario 2. Suppose that $\frac{p}{2\mu}\leq\alpha<\frac{p}{2(\mu-\sigma\sqrt{(1-\kappa)/\kappa})}$. Note that $(2\mu-\frac{p}{\alpha})\frac{p}{4\alpha}+\frac{p\mu}{2\alpha}=\frac{p}{\alpha}(\mu-\frac{p}{2\alpha})\leq\mu^2+\sigma^2$, where the inequality is due to the fact that $\frac{p}{\alpha}(\mu-\frac{p}{2\alpha})\leq\frac{\mu^2}{2}$. Hence, $\frac{p}{4\alpha}\leq\frac{\mu}{2}+\frac{\sigma^2}{2\mu-\frac{p}{\alpha}}$, yielding

$$L(q)=\begin{cases}\frac{\alpha}{2}\big(\frac{pq}{\alpha}+\mu^2+\sigma^2-\sqrt{(\frac{pq}{\alpha}+\mu^2+\sigma^2)^2-4\mu^2\frac{pq}{\alpha}}\big)-cq & q\leq\frac{\mu}{2}+\frac{\sigma^2}{2\mu-\frac{p}{\alpha}}\\ \frac{p}{2}\big(\mu-q-\frac{p}{4\alpha}-\sqrt{(q+\frac{p}{4\alpha}-\mu)^2+\sigma^2}\big)+(p-c)q & q\geq\frac{\mu}{2}+\frac{\sigma^2}{2\mu-\frac{p}{\alpha}}.\end{cases}\tag{EC.9}$$

Setting $L'(q)$ to 0 then yields $q^\star_\alpha=\big(\mu^2-\sigma^2+2f(1-\kappa)\mu\sigma\big)\cdot\frac{\alpha}{p}\leq\frac{\mu}{2}+\frac{\sigma^2}{2\mu-\frac{p}{\alpha}}$, where the inequality follows from the condition $\alpha<\frac{p}{2(\mu-\sigma\sqrt{(1-\kappa)/\kappa})}$.

Scenario 3. Suppose that $\alpha\geq\frac{p}{2(\mu-\sigma\sqrt{(1-\kappa)/\kappa})}$. Then $L(q)$ is given in (EC.9), which is concave. Setting $L'(q^\star_\alpha)=0$ yields $q^\star_\alpha=\mu+\sigma f(1-\kappa)-\frac{p}{4\alpha}\geq\frac{\mu}{2}+\frac{\sigma^2}{2\mu-\frac{p}{\alpha}}$.

By noting that *Scenario 1* and *Scenario 2* correspond to $\alpha<\frac{p}{2(\mu-\sigma\sqrt{(1-\kappa)/\kappa})}$ and *Scenario 3* corresponds to $\alpha\geq\frac{p}{2(\mu-\sigma\sqrt{(1-\kappa)/\kappa})}$, we then complete the proof. □

Proof of Proposition 2. When $p\geq\max\big\{\frac{\mu^2+\sigma^2}{\mu^2}c,2\alpha\mu\big\}$, it holds that $q^\star_\alpha=\big(\mu^2-\sigma^2+\mu\sigma\frac{1-2\kappa}{\sqrt{\kappa(1-\kappa)}}\big)\cdot\frac{\alpha}{p}=\frac{\alpha}{c}\cdot\frac{(\mu^2-\sigma^2+\mu\sigma(1-1/x))}{(x^2+1)}$, where we denote $x=\sqrt{\kappa/(1-\kappa)}$ to have the second equality. The derivative of $q^\star_\alpha$ with respect to x is $\frac{\partial q^\star_\alpha}{\partial x}=\frac{\alpha}{c}\cdot\frac{4\mu\sigma-\mu\sigma x^2-2x(\mu^2-\sigma^2)+\mu\sigma/x^2}{(x^2+1)^2}$. To determine the sign of $\frac{\partial q^\star_\alpha}{\partial x}$, it suffices to focus on the term $4\mu\sigma-\mu\sigma x^2-2x(\mu^2-\sigma^2)+\frac{\mu\sigma}{x^2}$. Note that when $x\geq1$ (*i.e.*, $\kappa>\frac{1}{2}$), we have $4\mu\sigma-\mu\sigma x^2-2x(\mu^2-\sigma^2)+\frac{\mu\sigma}{x^2}\leq 5\mu\sigma-\mu\sigma x^2-2x(\mu^2-\sigma^2)=-\mu\sigma(x+\frac{\mu^2-\sigma^2}{\mu\sigma})^2+5\mu\sigma-\frac{(\mu^2-\sigma^2)}{\mu\sigma}$. Since $\mu\sigma>0$, there must exist some $x_1\geq1$ such that when $x>x_1$, it holds that $4\mu\sigma-\mu\sigma x^2-2x(\mu^2-\sigma^2)+\frac{\mu\sigma}{x^2}\leq-\mu\sigma(x+\frac{\mu^2-\sigma^2}{\mu\sigma})^2+5\mu\sigma-\frac{(\mu^2-\sigma^2)}{\mu\sigma}<0$. As x is strictly increasing in (κ and) p, it is clear that $q^\star_\alpha$ is strictly decreasing in p when $p>p^\star_\alpha=\max\{\frac{\mu^2+\sigma^2}{\mu^2}c,2\alpha\mu,c(x_1^2+1)\}$, completing the proof. □

Proof of Proposition 3. The derivative of $q^\star_\alpha$ with respect to σ is

$$\frac{\partial q^\star_\alpha}{\partial\sigma}=\begin{cases}f(1-\kappa) & \sigma<(\mu-\frac{p}{2\alpha})\sqrt{\frac{\kappa}{1-\kappa}}\\ \big(-2\sigma+2\mu f(1-\kappa)\big)\cdot\frac{\alpha}{p} & (\mu-\frac{p}{2\alpha})\sqrt{\frac{\kappa}{1-\kappa}}\leq\sigma\leq\mu\sqrt{\frac{\kappa}{1-\kappa}},\end{cases}$$

which is decreasing in σ. When $\kappa\geq\frac{1}{2}$, $\frac{\partial q^\star_\alpha}{\partial\sigma}=f(1-\kappa)\geq0\ \ \forall\sigma\in\big[0,(\mu-\frac{p}{2\alpha})\sqrt{\frac{\kappa}{1-\kappa}}\big]$. We proceed by dividing the argument into two cases. On the one hand, if $\alpha\geq\frac{p-c}{\mu}$, we have $\mu f(1-\kappa)=\frac{\mu}{2}(\sqrt{\frac{\kappa}{1-\kappa}}-\sqrt{\frac{1-\kappa}{\kappa}})\leq(\mu-\frac{p}{2\alpha})\sqrt{\frac{\kappa}{1-\kappa}}$, which indicates that $\frac{\partial q^\star_\alpha}{\partial\sigma}=(-2\sigma+2\mu f(1-\kappa))\cdot\frac{\alpha}{p}\leq0\ \ \forall\sigma\in[(\mu-\frac{p}{2\alpha})\sqrt{\frac{\kappa}{1-\kappa}},\mu\sqrt{\frac{\kappa}{1-\kappa}}]$. On the other hand, if $\alpha<\frac{p-c}{\mu}$, it holds that $\mu f(1-\kappa)>(\mu-\frac{p}{2\alpha})\sqrt{\frac{\kappa}{1-\kappa}}$, which implies that $\frac{\partial q^\star_\alpha}{\partial\sigma}\geq0\ \ \forall\sigma\in[(\mu-\frac{p}{2\alpha})\sqrt{\frac{\kappa}{1-\kappa}},\mu f(1-\kappa)]$ and $\frac{\partial q^\star_\alpha}{\partial\sigma}\leq0\ \ \forall\sigma\in[\mu f(1-\kappa),\mu\sqrt{\frac{\kappa}{1-\kappa}}]$. Combining the two cases together then yields the desired result. □

Proof of Lemma 1. We proceed with the proof by dividing the argument into two scenarios.

Scenario 1.1. If $\frac{\hat{\mu}}{\hat{\sigma}}\geq\frac{\mu}{\sigma}$, by theorem 2.1 in Gelbrich (1990), for any $G\in\mathcal{A}_N$, it holds that $d(D,G)\geq(\hat{\mu}-\mu)^2+(\hat{\sigma}-\sigma)^2$, and hence $d(D,\mathcal{A}_N)\geq(\hat{\mu}-\mu)^2+(\hat{\sigma}-\sigma)^2$. In the following, we show that the inequality is tight. Consider the transformed random variable $\tilde{w}=\frac{\hat{\sigma}}{\sigma}\tilde{u}+\hat{\mu}-\frac{\hat{\sigma}\mu}{\sigma}\sim G^\dagger$ where $\tilde{u}\sim D$. Since the $\tilde{w}$ is a linear transformation of $\tilde{u}$, we have $d(D,G^\dagger)=(\hat{\mu}-\mu)^2+(\hat{\sigma}-\sigma)^2$ (Dowson and Landau 1982). Moreover, we can verify $\mathbb{E}_{G^\dagger}[\tilde{w}] = \hat{\mu}$, $\mathbb{E}_{G^\dagger}[\tilde{w}^2] = \hat{\mu}^2 + \hat{\sigma}^2$, and $G^\dagger\{\tilde{w} \in [0, +\infty)\} = 1$ (since $\frac{\hat{\mu}}{\hat{\sigma}} \ge \frac{\mu}{\sigma}$). Hence, $G^\dagger \in \mathcal{A}_N$, concluding $d(D, \mathcal{A}_N) = (\hat{\mu} - \mu)^2 + (\hat{\sigma} - \sigma)^2$.

Scenario 1.2. If $\frac{\hat{\mu}}{\hat{\sigma}} < \frac{\mu}{\sigma}$, we construct an upper bound for $d(D, \mathcal{A}_N)$. Consider the transformed random variable $\tilde{w} = k^\dagger \max\{0, \tilde{u} - t^\dagger\} \sim G^\dagger$ with $\tilde{u} \sim D$, where $k^\dagger > 0$ and $t^\dagger \ge 0$ satisfy

$$\int_{t^\dagger}^{+\infty} k^\dagger (u - t^\dagger) \mathrm{d}D(u) = \hat{\mu} \quad \text{and} \quad \int_{t^\dagger}^{+\infty} k^{\dagger 2} (u - t^\dagger)^2 \mathrm{d}D(u) = \hat{\mu}^2 + \hat{\sigma}^2. \tag{EC.10}$$

In the following, we first show that there exists $(k^\dagger, t^\dagger)$ satisfying (EC.10). Eliminating the variable k in (EC.10), it suffices to check whether there exists $t \ge 0$ such that

$$\hat{\mu}^2 \int_t^{+\infty} (u - t)^2 \mathrm{d}D(u) - (\hat{\mu}^2 + \hat{\sigma}^2) (\int_t^{+\infty} (u - t) \mathrm{d}D(u))^2 = 0.$$

Define $h(t) = \hat{\mu}^2 \int_t^{+\infty} (u - t)^2 \mathrm{d}D(u) - (\hat{\mu}^2 + \hat{\sigma}^2)(\int_t^{+\infty} (u - t) \mathrm{d}D(u))^2$. Setting the derivative $h'(t) = 2 \int_t^{+\infty} (u - t) \mathrm{d}D(u) ((\hat{\mu}^2 + \hat{\sigma}^2)(1 - D(t)) - \hat{\mu}^2)$ to 0 then yields $t^\diamond = D^{-1}(\frac{\hat{\sigma}^2}{\hat{\mu}^2 + \hat{\sigma}^2})$. It is straightforward to see that $h(0) = \hat{\mu}^2(\mu^2 + \sigma^2) - \mu^2(\hat{\mu}^2 + \hat{\sigma}^2) < 0$ and $\lim_{t \to +\infty} h(t) = 0$, which implies that $h(t^\diamond) > \lim_{t \to +\infty} h(t) = 0$ since $h(t)$ is decreasing in $(t^\diamond, +\infty)$. Therefore, there must exist some $t^\dagger \in [0, t^\diamond]$ such that $h(t^\dagger) = 0$, verifying the feasibility of (EC.10). This indicates that $\mathbb{E}_{G^\dagger}[\tilde{w}] = \hat{\mu}$ and $\mathbb{E}_{G^\dagger}[\tilde{w}^2] = \hat{\mu}^2 + \hat{\sigma}^2$. Additionally, since $G^\dagger\{\tilde{w} \in [0, +\infty)\} = 1$, it is immediate to see that $G^\dagger \in \mathcal{A}_N$. Subsequently, we identify the upper bounds for $k^\dagger$ and $t^\dagger$. Note that for any $t \in [0, t^\diamond]$, we have $h''(t) = 2\hat{\mu}^2(1 - D(t)) - 2d(t)(\hat{\mu}^2 + \hat{\sigma}^2) \int_t^{+\infty} (u - t) \mathrm{d}D(u) - 2(1 - D(t))^2 (\hat{\mu}^2 + \hat{\sigma}^2) \le 0$, where $d(t)$ is the density function of the distribution D. Hence, it holds that

$$\frac{h(t^\dagger) - h(0)}{t^\dagger - 0} \ge \frac{h(t^\diamond) - h(t^\dagger)}{t^\diamond - t^\dagger} \implies t^\dagger \le \frac{-h(0) t^\diamond}{h(t^\diamond) - h(0)}.$$

Note that as $N \to +\infty$, $h(0) \to 0$ and hence $t^\dagger \to 0$. This implies that for sufficiently large N, $t^\dagger \le \mu$. Since $\int_{t^\dagger}^{+\infty} (u - t^\dagger) \mathrm{d}D(u) = \mu - \int_0^{t^\dagger} u \mathrm{d}D(u) - t^\dagger(1 - D(t^\dagger)) \ge \mu - t^\dagger(1 - D(0))$, we then have

$$\frac{\hat{\mu}}{k^\dagger} \ge \mu - t^\dagger(1 - D(0)) \implies k^\dagger \le \frac{\hat{\mu}}{\mu - t^\dagger(1 - D(0))} \le \frac{\hat{\mu}}{\mu + \frac{h(0) t^\diamond}{h(t^\diamond) - h(0)}} \le \frac{\hat{\mu}}{\mu + \frac{h(0)}{h'(0)}},$$

where the last inequality follows from the fact that $\frac{h(t^\diamond) - h(0)}{t^\diamond} \le h'(0)$ since $h''(t) \le 0$ for any $t \in [0, t^\diamond]$. Plugging the expressions of $h(0)$ and $h'(0)$, it is then immediate to see that $k^\dagger \le \frac{\hat{\mu}}{\mu + \frac{h(0)}{h'(0)}} = \frac{2\mu^2 \hat{\sigma}^2}{\mu^2 \hat{\sigma}^2 + \hat{\mu}^2 \sigma^2}$. For the optimal-transport cost, note that the objective of the Kantorovich formulation as defined in (2) is no larger than that of the Monge formulation (Villani 2009), *i.e.*, $d(D, G^\dagger) \le \inf_{\psi: T_\psi[D] = G^\dagger} \int_0^{+\infty} (u - \psi(u))^2 \mathrm{d}D(u)$. Note that the function $\psi^\dagger(u) = k^\dagger \cdot \max\{u - t^\dagger, 0\}$ is feasible to the right-hand side problem. Hence,

$$\begin{aligned} d(D, \mathcal{A}_N) \le d(D, G^\dagger) &\le \int_0^t u^2 \mathrm{d}D(u) + \int_t^{+\infty} (k^\dagger(u - t^\dagger) - u)^2 \mathrm{d}D(u) \\ &= \mu^2 + \sigma^2 + \hat{\mu}^2 + \hat{\sigma}^2 - 2k^\dagger \int_{t^\dagger}^{+\infty} (u - t^\dagger) u \mathrm{d}D(u) \\ &\le \mu^2 + \sigma^2 + \hat{\mu}^2 + \hat{\sigma}^2 - \frac{(\mu^2 \hat{\sigma}^2 + \hat{\mu}^2 \sigma^2)(\hat{\mu}^2 + \hat{\sigma}^2)}{\mu^2 \hat{\sigma}^2} \\ &= (\mu - \hat{\mu})^2 + (\sigma - \hat{\sigma})^2 + 2\mu\hat{\mu} + 2\sigma\hat{\sigma} - \frac{(\mu^2 \hat{\sigma}^2 + \hat{\mu}^2 \sigma^2)(\hat{\mu}^2 + \hat{\sigma}^2)}{\mu^2 \hat{\sigma}^2}, \end{aligned}$$

where the third line follows from $u \ge u - t^\dagger$ and $k^\dagger \le \frac{2\mu^2 \hat{\sigma}^2}{\mu^2 \hat{\sigma}^2 + \hat{\mu}^2 \sigma^2}$. Besides,

$$2\mu\hat{\mu} + 2\sigma\hat{\sigma} - \frac{(\mu^2 \hat{\sigma}^2 + \hat{\mu}^2 \sigma^2)(\hat{\mu}^2 + \hat{\sigma}^2)}{\mu^2 \hat{\sigma}^2} = \frac{\hat{\mu}(\mu^2 \hat{\sigma}^2 - \hat{\mu}^2 \sigma^2)}{\mu \hat{\sigma}^2} + \frac{\sigma(\mu\hat{\sigma} - \hat{\mu}\sigma)}{\mu} + \frac{\hat{\sigma}(\hat{\mu}\sigma - \mu\hat{\sigma})}{\hat{\mu}} \le \frac{\mu^2 \hat{\sigma}^2 - \hat{\mu}^2 \sigma^2}{\sigma \hat{\sigma}},$$

where the inequality is due to $\frac{\hat{\mu}}{\hat{\sigma}} < \frac{\mu}{\sigma}$. Hence, we have $d(D, \mathcal{A}_N) \le (\mu - \hat{\mu})^2 + (\sigma - \hat{\sigma})^2 + \frac{\mu^2 \hat{\sigma}^2 - \hat{\mu}^2 \sigma^2}{\sigma \hat{\sigma}}$. Combining the results in these two scenarios then completes the proof. □

Proof of Proposition 4. We proceed in two steps. In the first step, we derive the concentration inequalities of the sample mean and variance, respectively. In the second step, we establish the concentration inequality for the mean-variance ambiguity set.

Step 1. Note that for any $x \in \mathbb{R}$, $\mathbb{E}_{D^N}[\exp(\frac{x}{N} \sum_{i=1}^N (\hat{v}_i - \mu))] = \prod_{i=1}^N \mathbb{E}_D[\exp(\frac{x}{N}(\hat{v}_i - \mu))] \le \mathbb{E}_D[\exp(-\frac{x^2 \nu^2}{2N})]$, where the expectation is taken with respect to the random sample $\hat{v}_i$, the equality follows from the fact that $\hat{v}_1, \ldots, \hat{v}_N$ are i.i.d, and the inequality follows from the fact that D is sub-Gaussian with variance proxy ν^2. Hence, $\hat{\mu} - \mu = \frac{1}{N} \sum_{i=1}^N (\hat{v}_i - \mu)$ is sub-Gaussian with variance proxy $\frac{\nu^2}{N}$. According to the concentration inequality of the sample mean for a sub-Gaussian distribution characterized in lemma 1.3 of Rigollet and Jan-Christian (2023), with probability at least $1 - \eta$, we have

$$|\hat{\mu} - \mu| \le \nu \sqrt{\frac{2 \log(2/\eta)}{N}}. \tag{EC.11}$$

By theorem 6.5 in Wainwright (2019), for any $\delta > 0$ there exist some constants C_1, C_2 and C_3 such that with probability at least $1-\eta$, it holds that

$$|\hat{\mu}^2 + \hat{\sigma}^2 - \mu^2 - \sigma^2| \leq \nu^2 C_1\big(\tfrac{1}{N} + \sqrt{\tfrac{1}{N}}\big) + \nu^2 \max\big\{\sqrt{\tfrac{\log(C_2/\eta)}{C_3 N}}, \tfrac{\log(C_2/\eta)}{C_3 N}\big\}.$$

Since $\frac{1}{N} \leq \frac{1}{\sqrt{N}}$ and $\sqrt{x} \leq 1 + x$ for any $x > 0$, with probability at least $1-\eta$, we further have

$$|\hat{\mu}^2 + \hat{\sigma}^2 - \mu^2 - \sigma^2| \leq \tfrac{\nu^2}{\sqrt{N}}\big(2C_1 + 1 + \tfrac{\log(C_2/\eta)}{C_3}\big). \tag{EC.12}$$

Note that

$$|\hat{\sigma}^2 - \sigma^2| \leq |\hat{\mu}^2 + \hat{\sigma}^2 - \mu^2 - \sigma^2| + |\hat{\mu}^2 - \mu^2| \leq |\hat{\mu}^2 + \hat{\sigma}^2 - \mu^2 - \sigma^2| + (\hat{\mu} - \mu)^2 + 2\mu|\hat{\mu} - \mu|, \tag{EC.13}$$

where the first inequality follows from the triangle inequality, and the second inequality follows from the fact that $|\hat{\mu}^2 - \mu^2| = |(\hat{\mu} - \mu)^2 + 2\mu(\mu - \hat{\mu})| \leq (\hat{\mu} - \mu)^2 + 2\mu|\hat{\mu} - \mu|$. Hence, it holds that

$$|\hat{\sigma} - \sigma| = \frac{|\hat{\sigma}^2 - \sigma^2|}{\hat{\sigma} + \sigma} \leq \frac{|\hat{\sigma}^2 - \sigma^2|}{\sigma} \leq \frac{|\hat{\mu}^2 + \hat{\sigma}^2 - \mu^2 - \sigma^2| + (\hat{\mu} - \mu)^2 + 2\mu|\hat{\mu} - \mu|}{\sigma}.$$

Applying the reverse union bound to inequalities (EC.11) and (EC.12), we then have with probability at least $1-\eta$,

$$|\hat{\sigma} - \sigma| \leq \tfrac{1}{\sigma}\Big(\tfrac{\nu^2}{\sqrt{N}}\big(2C_1 + 1 + \tfrac{\log(2C_2/\eta)}{C_3}\big) + 2\nu^2 \tfrac{\log(4/\eta)}{N} + 2\mu\nu\sqrt{\tfrac{2\log(4/\eta)}{N}}\Big) \leq \tfrac{\xi_1 + \xi_2 \log(1/\eta)}{\sqrt{N}}, \tag{EC.14}$$

where $\xi_1 = \frac{\nu}{\sigma}\big(\nu\big(2C_1 + 1 + \frac{\log(2C_2)}{C_3}\big) + 2(\nu + 2\mu)\log(4) + 2\mu\big)$ and $\xi_2 = \frac{\nu}{\sigma}\big(\frac{\nu}{C_3} + 2\nu + 4\mu\big)$. Here, the second inequality follows from the fact that $\frac{1}{N} \leq \frac{1}{\sqrt{N}}$ and $\sqrt{x} \leq 1 + x$ for any $x \geq 0$.

Step 2. To derive the concentration property for the mean-variance ambiguity, we divide the argument into two cases based on the expression of $d(D, \mathcal{A}_N)$.

Scenario 1. If $\frac{\hat{\mu}}{\hat{\sigma}} \geq \frac{\mu}{\sigma}$, according to Lemma 1, we have $d(D, \mathcal{A}_N) = (\hat{\mu} - \mu)^2 + (\hat{\sigma} - \sigma)^2$. By the concentration inequalities for sample mean and sample variance derived in equations (EC.11) and (EC.14), it is then immediate to see that with probability at least $1-\eta$, it holds that

$$(\hat{\mu} - \mu)^2 + (\hat{\sigma} - \sigma)^2 \leq \nu^2 \tfrac{4\log(2/\eta)}{N} + \tfrac{(\xi_1 + \xi_2 \log(1/\eta))^2}{N} \leq \tfrac{\xi_1^2 + 2\log(4)\nu^2 + (2\xi_1\xi_2 + 2\nu^2)\log(1/\eta) + \xi_2^2(\log(1/\eta))^2}{\sqrt{N}}. \tag{EC.15}$$

Let $\zeta_1 = \xi_1^2 + 2\log(4)\nu^2$ and $\zeta_2 = 2\xi_1\xi_2 + 2\nu^2$. When $\zeta_2^2 \geq 4\zeta_1\xi_2^2$, we have $\zeta_1 + \zeta_2\log(1/\eta) + \xi_2^2(\log(1/\eta))^2 \leq (\frac{\zeta_1}{\xi_2} + \xi_2\log(1/\eta))^2$. Setting $c_1 = \frac{\xi_1\xi_2 + \nu^2}{\xi_2}$ and $c_2 = \xi_2$ then yields the result. When $\zeta_2^2 < 4\zeta_1\xi_2^2$, we have $\zeta_1 + \zeta_2\log(1/\eta) + \xi_2^2(\log(1/\eta))^2 \leq (\sqrt{\zeta_1} + \xi_2\log(1/\eta))^2$. Setting $c_1 = \sqrt{\zeta_1}$ and $c_2 = \xi_2$ then yields the desired result.

Scenario 2. If $\frac{\hat{\mu}}{\hat{\sigma}} < \frac{\mu}{\sigma}$, it holds that

$$\tfrac{\mu^2\hat{\sigma}^2 - \hat{\mu}^2\sigma^2}{\sigma\hat{\sigma}} \leq \tfrac{\mu|\hat{\sigma}^2 - \sigma^2| + \sigma^2|\hat{\mu} - \mu|}{\sigma\hat{\sigma}} \leq \tfrac{\mu(|\hat{\mu}^2 + \hat{\sigma}^2 - \mu^2 - \sigma^2| + (\hat{\mu} - \mu)^2 + 2\mu|\hat{\mu} - \mu|) + \sigma^2|\hat{\mu} - \mu|}{\sigma\hat{\sigma}},$$

where the second inequality follows from inequality (EC.13). According to inequalities (EC.11) and (EC.12), with probability at least $1-\eta$, it holds that

$$\begin{aligned}
&\mu\big(|\hat{\mu}^2 + \hat{\sigma}^2 - \mu^2 - \sigma^2| + (\hat{\mu} - \mu)^2 + 2\mu|\hat{\mu} - \mu|\big) + \sigma^2|\hat{\mu} - \mu| \\
\leq\ & \tfrac{\nu^2\mu}{\sqrt{N}}\big(2C_1 + 1 + \tfrac{\log(2C_2/\eta)}{C_3}\big) + (2\mu^2 + \sigma^2)\nu\sqrt{\tfrac{2\log(4/\eta)}{N}} + \mu\nu^2\tfrac{2\log(4/\eta)}{N} \\
\leq\ & \tfrac{\delta_1 + \delta_2\log(1/\eta)}{\sqrt{N}},
\end{aligned}$$

where $\delta_1 = \nu^2\mu\big(2C_1 + 1 + \frac{\log(2C_2)}{C_3}\big) + 2\log(4)(2\mu^2 + \sigma^2 + \nu\mu)\nu + (2\mu^2 + \sigma^2)\nu$ and $\delta_2 = \frac{\nu^2\mu}{C_3} + 2(2\mu^2 + \sigma^2)\nu + 2\mu\nu^2$. When N is sufficiently large, $\sigma > \frac{\xi_1 + \xi_2\log(1/\eta)}{\sqrt{N}}$, and then (EC.14) implies that $\hat{\sigma} \leq \sigma - \frac{\xi_1 + \xi_2\log(1/\eta)}{\sqrt{N}}$ with probability at least $1-\eta$. Hence, we have

$$\tfrac{\mu^2\hat{\sigma}^2 - \hat{\mu}^2\sigma^2}{\sigma\hat{\sigma}} \leq \frac{\frac{\delta_1 + \delta_2\log(1/\eta)}{\sqrt{N}}}{\sigma(\sigma - \frac{\xi_1 + \xi_2\log(1/\eta)}{\sqrt{N}})} = \tfrac{\delta_1 + \delta_2\log(1/\eta)}{\sqrt{N}\sigma^2 - \sigma(\xi_1 + \xi_2\log(1/\eta))},$$

holds with probability at least $1-\eta$. Note that for sufficiently large N, we have $\sqrt{N}\sigma^2 > \delta_1 + \delta_2\log(1/\eta) + \sigma(\xi_1 + \xi_2\log(1/\eta))$, which implies that $\frac{\delta_1 + \delta_2\log(1/\eta)}{\sqrt{N}\sigma^2 - \sigma(\xi_1 + \xi_2\log(1/\eta))} < \frac{\delta_1 + \delta_2\log(1/\eta) + \sigma(\xi_1 + \xi_2\log(1/\eta))}{\sqrt{N}\sigma^2}$. This inequality, together with (EC.15), implies that with probability at least $1-\eta$, for sufficiently large N, it holds that

$$d(D, \mathcal{A}_N) \leq (\hat{\mu} - \mu)^2 + (\hat{\sigma} - \sigma)^2 + \tfrac{\mu^2\hat{\sigma}^2 - \hat{\mu}^2\sigma^2}{\sigma\hat{\sigma}} \leq \tfrac{(\xi_1^2 + 2\log(2)\nu^2)\sigma^2 + \delta_1 + \sigma\xi_1 + (2\xi_1\xi_2\sigma^2 + 2\nu^2\sigma^2 + \delta_2 + \sigma\xi_2)\log(1/\eta) + \xi_2^2\sigma^2(\log(1/\eta))^2}{\sqrt{N}\sigma^2}.$$

Using a similar argument as in *Scenario 1*, we can also show that with probability at least $1-\eta$, for sufficiently large N, $d(D,\mathcal{A}_N) \le \frac{(c_1+c_2\log(1/\eta))^2}{\sqrt{N}}$ where c_1 and c_2 are some constants that only depend on μ, σ and ν. $\square$

Proof of Theorem 3. We proceed in two steps. In the first step, leveraging the equivalence between

$$\Upsilon^\star_\varepsilon = \max_{q\ge 0} \min_{d(F,\mathcal{A})\le\varepsilon} \mathbb{E}_F[\pi(q,\tilde{u})] \tag{EC.16}$$

and MISSPECIFICATION (as shown in Lemma EC.2), we characterize the relationship between ε and α. In the second step, we translate the finite-sample performance guarantee of (EC.16) characterized in Proposition 4 as the performance guarantee of MISSPECIFICATION.

Step 1. Given N and $\varepsilon_N + d(F,D)$, consider the constrained problem (EC.16) with $\mathcal{A}=\mathcal{A}_N$ and $\varepsilon = \varepsilon_N + d(F,D)$. Suppose that α_N is an index of misspecification aversion such that MISSPECIFICATION with $\mathcal{A}=\mathcal{A}_N$ and $\alpha=\alpha_N$ has the same optimal solution as that of the corresponding constrained problem (EC.16). Denote by $\Pi^\star_{\alpha_N}$ the optimal value of MISSPECIFICATION with $\alpha=\alpha_N$. By Lemma EC.2, we have

$$\Upsilon^\star_{\varepsilon_N+d(F,D)} = \max_{\alpha\ge 0}\left\{\Pi^\star_\alpha - (\varepsilon_N + d(F,D))\alpha\right\} = \Pi^\star_{\alpha_N} - (\varepsilon_N+d(F,D))\alpha_N.$$

In the following, we characterize the expression of α_N. Plugging the expression of the optimal order quantity $q^\star_\alpha$ into the worst-case transformed expectation characterized in Proposition 1, we have

$$\Pi^\star_\alpha - (\varepsilon_N+d(F,D))\alpha = \begin{cases} \kappa\alpha\hat{v}^{\star 2} - (\varepsilon_N + d(F,D))\alpha & \alpha < \frac{p}{2\hat{v}^\star} \\ (p-c)\hat{v}^\star - \frac{p(p-c)}{4\alpha} - (\varepsilon_N+d(F,D))\alpha & \alpha \ge \frac{p}{2\hat{v}^\star}, \end{cases}$$

where $\hat{v}^\star = \hat{\mu} - \hat{\sigma}\sqrt{\frac{1-\kappa}{\kappa}}$. Note that $\Pi^\star_\alpha - (\varepsilon_N + d(F,D))\alpha$ is concave in α. If $\varepsilon_N + d(F,D) \ge \kappa\hat{v}^{\star 2}$, then $\Pi^\star_\alpha - (\varepsilon_N+d(F,D))\alpha$ is decreasing in α, and hence $\max_{\alpha\ge 0}\{\Pi^\star_\alpha - (\varepsilon_N + d(F,D)))\alpha\} = 0$ with $\alpha_N = 0$. If $\varepsilon_N + d(F,D) < \kappa\hat{v}^{\star 2}$, then by the first-order optimality condition, the maximum is attained at $\alpha_N = \frac{\sqrt{p(p-c)}}{2\sqrt{\varepsilon_N+d(F,D)}}$ and $\max_{\alpha\ge 0}\{\Pi^\star_\alpha - (\varepsilon_N+d(F,D))\alpha\} = \Pi^\star_{\alpha_N} - \frac{1}{2}\sqrt{p(p-c)(\varepsilon_N+d(F,D))}$. To summarize the second step, we have obtained

$$\alpha_N = \begin{cases} \frac{\sqrt{p(p-c)}}{2\sqrt{\varepsilon_N+d(F,D)}} & \varepsilon_N + d(F,D) < \kappa\hat{v}^{\star 2} \\ 0 & \varepsilon_N+d(F,D) \ge \kappa\hat{v}^{\star 2} \end{cases} \quad \text{and} \quad \Upsilon^\star_{\varepsilon_N+d(F,D)} = \left(\Pi^\star_{\alpha_N} - \tfrac{1}{2}\sqrt{p(p-c)(\varepsilon_N+d(F,D))}\right)^+.$$

Step 2. Denote by $q^\star_{\alpha_N}$ the optimal solution to MISSPECIFICATION with $\alpha=\alpha_N$, which, by Lemma EC.2, is also the optimal solution to the constrained problem (EC.16). Note that the empirical distribution $\hat{G} = \frac{1}{N}\sum_{i=1}^N \hat{v}_i$ satisfies $\mathbb{E}_{\hat{G}}[\tilde{v}] = \hat{\mu}$ and $\mathbb{E}_{\hat{G}}[\tilde{v}^2] = \hat{\mu}^2 + \hat{\sigma}^2$. Hence, $\hat{G}\in\mathcal{A}_N$. For any F such that $d(F,\mathcal{A}_N) \le \varepsilon_N + d(F,D)$, we have $\mathbb{E}_F[\pi(q^\star_{\alpha_N},\tilde{v})] \ge \Upsilon^\star_{\varepsilon_N+d(F,D)}$. Therefore,

$$\mathbb{P}_{D^N}\left\{\mathbb{E}_F[\pi(q^\star_{\alpha_N},\tilde{v})] \ge \Upsilon^\star_{\varepsilon_N+d(F,D)}\right\} \ge \mathbb{P}_{D^N}\{d(F,\mathcal{A}_N)\le \varepsilon_N + d(F,D)\} \ge \mathbb{P}_{D^N}\{d(F,\hat{G}) \le \varepsilon_N + d(F,D)\},$$

where the second inequality follows since $d(D,\mathcal{A}_N) \le \varepsilon_N$ implies $d(F,\mathcal{A}_N) \le d(D,\mathcal{A}_N) + d(F,D) \le \varepsilon_N + d(F,D)$. This, together with the inequality in Proposition 4 implies that $\mathbb{P}_{D^N}\{\mathbb{E}_F[\pi(q^\star_{\alpha_N},\tilde{u})] \ge \Upsilon^\star_{\varepsilon_N+d(F,D)}\} \ge 1-\eta$. Plugging the expressions of α_N and $\Upsilon^\star_{\varepsilon_N+d(F,D)}$ given in the first step, we obtain the desired result. $\square$

Proof of Theorem 4. For ease of notation, define $\mathcal{G} = \{G\in\mathcal{P}_M \mid \mathbb{E}_G[\tilde{v}_i] = \mu_i \quad \forall i\in[M]\}$ and $\mathcal{C}_i = \{G \in \mathcal{P} \mid \mathbb{E}_G[\tilde{v}_i] = \mu_i\}$ for $i\in[M]$. Introducing the dual variable $\lambda\ge 0$ to the sum-of-variance constraint in the ambiguity set, then the MULTIPLE model can be equivalently reformulated as

$$\max_{\lambda\ge 0}\left\{-\lambda K + \max_{\boldsymbol{q}\ge 0}\min_{F\in\mathcal{P}_M,\, G\in\mathcal{G}} \mathbb{E}_F\Big[\sum_{i\in[M]}\pi(q_i,\tilde{u}_i)\Big] + \lambda\cdot\mathbb{E}_G\Big[\sum_{i\in[M]}\tilde{v}_i^2\Big] + \alpha\cdot d(F,G)\right\},$$

which, by noting that $\mathcal{G}$ is decomposable with respect to multiple products, further reduces to

$$\max_{\lambda\ge 0}\left\{-\lambda K + \sum_{i\in[M]}\max_{q_i\ge 0}\min_{F_i\in\mathcal{P},\, G_i\in\mathcal{C}_i}\left\{\mathbb{E}_{F_i}[\pi(q_i,\tilde{u}_i)] + \lambda\cdot\mathbb{E}_{G_i}[\tilde{v}_i^2] + \alpha\cdot d(F_i,G_i)\right\}\right\}. \tag{EC.17}$$

In the remainder of the proof, we solve for the optimal $\lambda^\star$ and $\boldsymbol{q}^\star_\alpha$ of problem (EC.17). Invoking the interchangeability principle characterized in Lemma EC.3, then problem (EC.17) becomes

$$\max_{\lambda\ge 0}\left\{-\lambda K + \sum_{i\in[M]}\max_{q_i\ge 0}\min_{G_i\in\mathcal{C}_i}\mathbb{E}_{G_i}\Big[\min_{u_i\ge 0}\{\pi(q_i,u_i) + \lambda\cdot\tilde{v}_i^2 + \alpha\cdot(u_i-\tilde{v}_i)^2\}\Big]\right\},$$

which, by defining $\digamma_i(\lambda, q_i, v_i) = \lambda \cdot v_i^2 + \min_{u_i \geq 0}\{\pi(q_i, u_i) + \alpha \cdot (u_i - v_i)^2\}$ (see Lemma EC.4 for its closed-form expression) for each $i \in [M]$ and $u_i \geq 0$, can be equivalently written as

$$\max_{\lambda \geq 0} \Big\{ -\lambda K + \sum_{i\in[M]} \max_{q_i \geq 0} \min_{G_i \in \mathcal{C}_i} \mathbb{E}_{G_i}[\digamma_i(\lambda, q_i, \tilde{v}_i)] \Big\}. \tag{EC.18}$$

In the following, our remaining proof proceeds in three steps: deriving the expression for $L_i(q_i) = \min_{G_i\in\mathcal{C}_i} \mathbb{E}_{G_i}[\digamma_i(\lambda, q_i, \tilde{v}_i)]$ (*Step* 1), optimizing over q_i to solve $\max_{q_i\geq 0} L_i(q_i)$ for each $i \in [M]$ (*Step* 2), and finally, optimizing over $\lambda \geq 0$ (*Step* 3). Note that given $\lambda \geq 0$, we solve the inner maximization of problem (EC.18) over $q_i \geq 0$ for each $i \in [M]$ in *Step* 1 and *Step* 2.

Step 1. We drop the subscript 'i' to avoid clutter. Given $q \geq 0$, $L(q)$ is a classical moment problem:

$$\begin{aligned} \min_{G}\ & \int_{\mathbb{R}_+} \digamma(\lambda, q, v)\, \mathrm{d}G(v) && \\ \text{s.t.}\ & \int_{\mathbb{R}_+} v\, \mathrm{d}G(v) = \mu && \cdots\ s_\alpha \\ & \int_{\mathbb{R}_+} \mathrm{d}G(v) = 1 && \cdots\ t_\alpha \\ & G \in \mathcal{M}_+, && \end{aligned} \tag{Primal}$$

whose dual is given by

$$\begin{aligned} \max_{s_\alpha,\, t_\alpha}\ & \mu s_\alpha - t_\alpha \\ \text{s.t.}\ & v s_\alpha - t_\alpha \leq \digamma(\lambda, q, v) \quad \forall v \geq 0. \end{aligned} \tag{Dual}$$

We next derive the expression of $L(q)$ by constructing a pair of primal and dual feasible solutions that attain the same objective value (that is, strong duality holds between Primal and Dual). The argument breaks into nine scenarios based on the value of q.

Scenario 1.1. When $q \leq \frac{p}{4\alpha}$ and $\frac{\alpha\lambda\mu^2}{p(\lambda+\alpha)} \leq q \leq \frac{\alpha(\lambda+\alpha)\mu^2}{p\lambda}$, we first construct a feasible distribution to Primal as follows:

$$G_\alpha = \left(\frac{\sqrt{\frac{\lambda+\alpha}{\lambda}} - \mu\sqrt{\frac{\alpha}{pq}}}{\sqrt{\frac{\lambda+\alpha}{\lambda}} - \sqrt{\frac{\lambda}{\lambda+\alpha}}} \right) \cdot \delta_{\sqrt{\frac{\lambda pq}{\alpha(\lambda+\alpha)}}} + \left(\frac{\mu\sqrt{\frac{\alpha}{pq}} - \sqrt{\frac{\lambda}{\lambda+\alpha}}}{\sqrt{\frac{\lambda+\alpha}{\lambda}} - \sqrt{\frac{\lambda}{\lambda+\alpha}}} \right) \cdot \delta_{\sqrt{\frac{(\lambda+\alpha)pq}{\alpha\lambda}}}. \tag{EC.19}$$

One can verify that $G_\alpha \in \mathcal{A}$ and the corresponding primal objective value is equal to $\mathbb{E}_{G_\alpha}[\digamma(\lambda, q, \tilde{v})] = 2\mu\sqrt{\lambda(\lambda+\alpha)pq/\alpha} - \frac{\lambda pq}{\alpha} - cq$. We next construct a dual feasible that attains the same dual objective value. Consider the following solution

$$s_\alpha = 2\sqrt{\frac{\lambda(\lambda+\alpha)pq}{\alpha}}, \quad t_\alpha = \frac{\lambda pq}{\alpha} + cq, \tag{EC.20}$$

which satisfies $s_\alpha\mu - t_\alpha = 2\mu\sqrt{\lambda(\lambda+\alpha)pq/\alpha} - \frac{\lambda pq}{\alpha} - cq$. Similarly to the proof of Proposition 1, one can verify that solution (EC.20) is feasible to Dual and this establishes the strong duality.

Scenario 1.2. When $q \leq \frac{p}{4\alpha}$ and $q \leq \frac{\alpha\lambda\mu^2}{p(\lambda+\alpha)}$, we construct a primal feasible solution $G_\alpha = \delta_\mu$ with a primal objective value $\mathbb{E}_{G_\alpha}[\digamma(\lambda, q, \tilde{v})] = \lambda\mu^2 + pq - cq$. Consider the solution $s_\alpha = 2\lambda\mu$ and $t_\alpha = \lambda\mu^2 - (p-c)q$, which is feasible and satisfies $s_\alpha\mu - t_\alpha = \lambda\mu^2 + pq - cq$.

Scenario 1.3. When $q \leq \frac{p}{4\alpha}$ and $q \geq \frac{\alpha(\lambda+\alpha)\mu^2}{p\lambda}$, we construct a primal feasible solution $G_\alpha = \delta_\mu$ with a primal objective value $\mathbb{E}_{G_\alpha}[\digamma(\lambda, q, \tilde{v})] = (\lambda+\alpha)\mu^2 - cq$. Consider the solution $s_\alpha = 2(\lambda+\alpha)\mu$ and $t_\alpha = (\lambda+\alpha)\mu^2 + cq$, which is feasible and satisfies $s_\alpha\mu - t_\alpha = (\lambda+\alpha)\mu^2 - cq$.

Scenario 1.4. When $\frac{p}{4\alpha} \leq q \leq \frac{p}{4\alpha} + \frac{p}{4\lambda}$ and $\frac{\alpha\lambda\mu^2}{p(\lambda+\alpha)} \leq q \leq \frac{\alpha(\lambda+\alpha)\mu^2}{p\lambda}$, we consider the pair of primal feasible solution (EC.19) and dual solution (EC.20). Upon the results established in *Scenario* 1.1, we can verify that solution (EC.6) is feasible to Dual.

Scenario 1.5. When $\frac{p}{4\alpha} \leq q \leq \frac{p}{4\alpha} + \frac{p}{4\lambda}$ and $q \leq \frac{\alpha\lambda\mu^2}{p(\lambda+\alpha)}$, we construct a primal feasible solution $G_\alpha = \delta_\mu$ with a primal objective value $\mathbb{E}_{G_\alpha}[\digamma(\lambda, q, \tilde{v})] = \lambda\mu^2 + pq - cq$. Consider the solution $s_\alpha = 2\lambda\mu$ and $t_\alpha = \lambda\mu^2 - (p-c)q$, which is feasible and satisfies $s_\alpha\mu - t_\alpha = \lambda\mu^2 + pq - cq$.

Scenario 1.6. When $\frac{p}{4\alpha} \leq q \leq \frac{p}{4\alpha} + \frac{p}{4\lambda}$ and $q \geq \frac{\alpha(\lambda+\alpha)\mu^2}{p\lambda}$, we construct a primal feasible solution $G_\alpha = \delta_\mu$ with a primal objective value $\mathbb{E}_{G_\alpha}[\digamma(\lambda, q, \tilde{v})] = (\lambda+\alpha)\mu^2 - cq$. Consider the solution $s_\alpha = 2(\lambda+\alpha)\mu$ and $t_\alpha = (\lambda+\alpha)\mu^2 + cq$, which is feasible and satisfies $s_\alpha\mu - t_\alpha = (\lambda+\alpha)\mu^2 - cq$.

Scenario 1.7. When $q \geq \frac{p}{4\alpha} + \frac{p}{4\lambda}$ and $\mu - \frac{p}{4\alpha} - \frac{p}{4\lambda} \leq q \leq \mu - \frac{p}{4\alpha} + \frac{p}{4\lambda}$, we first construct a primal feasible solution: $G_\alpha = \big(\frac{1}{2} - \frac{2\lambda}{p}(\mu - q - \frac{p}{4\alpha})\big) \cdot \delta_{q+\frac{p}{4\alpha}-\frac{p}{4\lambda}} + \big(\frac{1}{2} + \frac{2\lambda}{p}(\mu - q - \frac{p}{4\alpha})\big) \cdot \delta_{q+\frac{p}{4\alpha}+\frac{p}{4\lambda}}$. One can verify that $G_\alpha \in \mathcal{A}$ and the corresponding objective value under G_α is equal to $\mathbb{E}_{G_\alpha}[\digamma(\lambda, q, \tilde{v})] = \big(2\lambda q + \frac{(\lambda+\alpha)p}{2\alpha}\big)\mu - \lambda q^2 -$

$\frac{(\lambda+\alpha)pq}{2\alpha} - \frac{(\lambda+\alpha)^2p^2}{16\alpha^2\lambda} + (p-c)q$. Moreover, we consider the solution: $s_\alpha = 2\lambda q + \frac{(\lambda+\alpha)p}{2\alpha}$, $t_\alpha = \lambda q^2 + \frac{(\lambda+\alpha)pq}{2\alpha} + \frac{(\lambda+\alpha)^2p^2}{16\alpha^2\lambda} - (p-c)q$, which is dual feasible.

Scenario 1.8. When $q \geq \frac{p}{4\alpha}$ and $q \leq \mu - \frac{p}{4\alpha} - \frac{p}{4\lambda}$, we construct a primal feasible solution $G_\alpha = \delta_\mu$ with a primal objective value $\mathbb{E}_{G_\alpha}[\digamma(\lambda, q, \tilde{v})] = \lambda\mu^2 + pq - cq$. Consider the solution $s_\alpha = 2\lambda\mu$ and $t_\alpha = \lambda\mu^2 - (p-c)q$, which satisfies $s_\alpha\mu - t_\alpha = \lambda\mu^2 + pq - cq$. The feasibility of (s_α, t_α) can be verified easily.

Scenario 1.9. When $q \geq \frac{p}{4\alpha}$ and $q \geq \mu - \frac{p}{4\alpha} + \frac{p}{4\lambda}$, we first construct a primal feasible solution $G_\alpha = \delta_\mu$ with a primal objective value $\mathbb{E}_{G_\alpha}[\digamma(\lambda, q, \tilde{v})] = \lambda\mu^2 + p(\mu - \frac{p}{4\alpha}) - cq$. Consider the solution $s_\alpha = 2\lambda\mu + p$ and $t_\alpha = \lambda\mu^2 + \frac{p^2}{4\alpha} + cq$, which is feasible and satisfies $s_\alpha\mu - t_\alpha = \lambda\mu^2 + p(\mu - \frac{p}{4\alpha}) - cq$.

To summarize *Step 1*, we note that when the constructed primal feasible distribution is $G_\alpha = \delta_\mu$ (*i.e.*, *Scenarios 1.2, 1.3, 1.5, 1.6, 1.8, 1.9*), the objective function $L(q)$ is either increasing or decreasing in q, implying that the maximum can not be attained in these scenarios. Therefore, we only need to focus on the remaining scenarios (*i.e.*, *Scenarios 1.1, 1.4, 1.7*) where

$$L(q) = \begin{cases} 2\mu\sqrt{\frac{\lambda(\lambda+\alpha)pq}{\alpha}} - \frac{\lambda pq}{\alpha} - cq & q \in \mathcal{Q}_1 \\ \left(2\lambda q + \frac{(\lambda+\alpha)p}{2\alpha}\right)\mu - \lambda q^2 - \frac{(\lambda+\alpha)pq}{2\alpha} - \frac{(\lambda+\alpha)^2p^2}{16\alpha^2\lambda} + (p-c)q & q \in \mathcal{Q}_2 \end{cases}$$

with $\mathcal{Q}_1 = \{q \mid q \leq \frac{p}{4\alpha} + \frac{p}{4\lambda},\ \frac{\alpha\lambda\mu^2}{p(\lambda+\alpha)} \leq q \leq \frac{\alpha(\lambda+\alpha)\mu^2}{p\lambda}\}$ and $\mathcal{Q}_2 = \{q \mid q \geq \frac{p}{4\alpha} + \frac{p}{4\lambda},\ \mu - \frac{p}{4\alpha} - \frac{p}{4\lambda} \leq q \leq \mu - \frac{p}{4\alpha} + \frac{p}{4\lambda}\}$.

Step 2. We consider three scenarios based on the values of α and λ to solve for the optimal q.

Scenario 2.1. Suppose that $\lambda \leq \frac{c}{2\mu}$. For any $q \in \mathcal{Q}_1$, setting the derivative of $L(q)$ to 0 yields

$$q^\star_\alpha(\lambda) = \frac{\lambda(\lambda+\alpha)p\mu^2}{\alpha(\lambda p/\alpha + c)^2}. \tag{EC.21}$$

One can verify that $\frac{\alpha\lambda\mu^2}{p(\lambda+\alpha)} \leq q^\star_\alpha \leq \frac{\alpha(\lambda+\alpha)\mu^2}{p\lambda}$. We next show that $q^\star_\alpha(\lambda) \leq \frac{p}{4\lambda} + \frac{p}{4\alpha}$. Since given $\alpha \geq 0$, $\frac{\partial}{\partial\alpha}\left(q^\star_\alpha(\lambda) - \frac{p}{4\alpha}\right) = \frac{p\lambda^2\mu^2(\alpha(p-c)+\alpha p+\lambda p)}{(\alpha c+\lambda p)^3} + \frac{p}{4\alpha^2} \geq 0$, and hence we have $q^\star_\alpha(\lambda) - \frac{p}{4\alpha} \leq \lim_{\alpha\to\infty}\left(q^\star_\alpha(\lambda) - \frac{p}{4\alpha}\right) = \frac{\lambda p\mu^2}{c^2} \leq \frac{p}{4\lambda}$, where the last inequality follows from $\lambda \leq \frac{c}{2\mu}$. Since $L(q)$ is concave, $q^\star_\alpha(\lambda)$ is indeed optimal.

Scenario 2.2. Suppose that $\lambda \geq \frac{c}{2\mu}$ and $\alpha < \frac{p}{2(\mu - c/(2\lambda))}$. For $q \in \mathcal{Q}_1$, setting $L'(q)$ to 0 yields $q^\star_\alpha(\lambda) = \frac{\lambda(\lambda+\alpha)p\mu^2}{\alpha(\lambda p/\alpha+c)^2} \in [\frac{\alpha\lambda\mu^2}{p(\lambda+\alpha)}, \frac{\alpha(\lambda+\alpha)\mu^2}{p\lambda}]$. Since $q^\star_\alpha(\lambda) - \frac{p}{4\alpha}$ is increasing in α, we have $q^\star_\alpha(\lambda) - \frac{p}{4\alpha} < \lim_{\alpha\to\frac{p}{2(\mu-c/(2\lambda))}}\left(q^\star_\alpha(\lambda) - \frac{p}{4\alpha}\right) = \frac{p}{4\lambda}$. This implies that $q^\star_\alpha(\lambda) \in \mathcal{Q}_1$. Hence, $q^\star_\alpha(\lambda)$ is optimal.

Scenario 2.3. Suppose that $\lambda \geq \frac{c}{2\mu}$ and $\alpha \geq \frac{p}{2(\mu - c/(2\lambda))}$. For $q \in \mathcal{Q}_2$, setting $L'(q)$ to 0 yields

$$q^\star_\alpha(\lambda) = \mu + \frac{p-2c}{4\lambda} - \frac{p}{4\alpha}. \tag{EC.22}$$

One can verify that $\mu - \frac{p}{4\alpha} - \frac{p}{4\lambda} \leq q^\star_\alpha(\lambda) \leq \mu - \frac{p}{4\alpha} + \frac{p}{4\lambda}$. Since $\lambda \geq \frac{c}{2\mu}$ and $\alpha \geq \frac{p}{2(\mu-c/(2\lambda))}$, we have $q^\star_\alpha(\lambda) = \mu + \frac{p-2c}{4\lambda} - \frac{p}{2\alpha} + \frac{p}{4\alpha} \geq \frac{p}{4\alpha} + \frac{p}{4\lambda}$, which indicates that $q^\star_\alpha(\lambda) \in \mathcal{Q}_2$. Therefore, $q^\star_\alpha(\lambda)$ is optimal.

To summarize *Step 2*, we note that *Scenario 2.1* and *Scenario 2.2* correspond to $\alpha < \frac{p}{2(\mu - c/(2\lambda))^+}$ and *Scenario 2.3* corresponds to $\alpha \geq \frac{p}{2(\mu - c/(2\lambda))^+}$. Hence, we have

$$q^\star_\alpha = \begin{cases} \mu + \frac{p-2c}{4\lambda} - \frac{p}{4\alpha} & \alpha \geq \frac{p}{2(\mu-c/(2\lambda))^+} \quad \left(\lambda \geq \frac{c}{(2\mu - p/\alpha)^+}\right) \\ \frac{\lambda(\lambda+\alpha)p\mu^2}{\alpha(\lambda p/\alpha+c)^2} & \alpha < \frac{p}{2(\mu-c/(2\lambda))^+} \quad \left(\lambda < \frac{c}{(2\mu - p/\alpha)^+}\right). \end{cases}$$

Step 3. We next solve the optimal $\lambda^\star$ of the outer maximization of problem (EC.18), whose objective function we denote by

$$Q(\lambda) = -\lambda K + \sum_{i\in[M]} \max_{q_i\geq 0} \min_{G_i\in\mathcal{C}_i} \mathbb{E}_{G_i}[\digamma_i(\lambda, q_i, \tilde{v}_i)] = -\lambda K + \sum_{i\in[M]} L_i(q^\star_{i,\alpha}(\lambda)).$$

Note that given $v_i \geq 0$ and $i \in [M]$, $\digamma_i(\lambda, q_i, v_i)$ is jointly concave in (λ, q_i). Since concavity is preserved under expectation and maximization, $Q(\lambda)$ is concave in λ. When $\lambda < \bar{\lambda}_i = \frac{c_i}{(2\mu_i - p_i/\alpha)^+}$ (*Scenario 2.1* and *Scenario 2.2*), plugging the expression of $q^\star_{i,\alpha}(\lambda)$ in (EC.21), we have $L_i(q^\star_{i,\alpha}(\lambda)) = \frac{\lambda p_i\mu_i^2\alpha(\lambda+\alpha)}{\lambda p_i + \alpha c_i}$ and $\frac{\partial L_i(q^\star_{i,\alpha}(\lambda))}{\partial\lambda} = \frac{\mu_i^2 p_i(\alpha^2 c_i + 2\alpha c_i\lambda + \lambda^2 p_i)}{(\lambda p_i + \alpha c_i)^2}$. When $\lambda \geq \bar{\lambda}_i = \frac{c_i}{(2\mu_i - p_i/\alpha)^+}$ (*Scenario 2.3*), plugging the expression of $q^\star_{i,\alpha}(\lambda)$ in (EC.22), we have $L_i(q^\star_{i,\alpha}(\lambda)) = \frac{\lambda(c_i-p_i)p_i + \alpha(c^2 + 4\lambda\mu_i(\mu_i+p_i) - c(4\lambda\mu_i + p_i))}{4\alpha\lambda}$ and $\frac{\partial L_i(q^\star_{i,\alpha}(\lambda))}{\partial\lambda} = \frac{c_i(p_i - c_i)}{4\lambda^2}$. For any $\lambda \in (\bar{\lambda}_{j-1}, \bar{\lambda}_j)$ with $j \in [M+1]$, we then have

$$\begin{aligned} Q(\lambda) &= -\lambda K + \sum_{i\in[M]\setminus[j-1]} \frac{\lambda p_i\mu_i^2\alpha(\lambda+\alpha)}{\lambda p_i + \alpha c_i} + \sum_{i\in[j-1]} \frac{4\alpha\lambda\mu_i(\mu_i + p_i - c_i) - (\lambda p_i + \alpha c_i)(p_i - c_i)}{4\alpha\lambda} \\ Q'(\lambda) &= -K + \sum_{i\in[M]\setminus[j-1]} \frac{\mu_i^2 p_i(\alpha^2 c_i + 2\alpha c_i\lambda + \lambda^2 p_i)}{(\lambda p_i + \alpha c_i)^2} + \sum_{i\in[j-1]} \frac{c_i(p_i - c_i)}{4\lambda^2} = -K + \Theta_j(\lambda). \end{aligned}$$

It can be noted that $Q(\lambda)$ is concave and $Q'(\lambda)$ is always decreasing in $(\bar{\lambda}_{j-1}, \bar{\lambda}_j)$, where at the end points it holds that $Q'_-(\bar{\lambda}_j) \geq Q'_+(\bar{\lambda}_j)$ for any $j \in [M+1]$. On the one hand, if $Q'_+(\bar{\lambda}_{i^\star-1}) = -K + \Theta_{i^\star}(\bar{\lambda}_{i^\star-1}) \leq 0$, then when $i^\star = 1$, we have $Q'(\lambda) \leq Q'_+(\bar{\lambda}_0) \leq 0$ for $\lambda \geq 0$, and hence $\lambda^\star = \bar{\lambda}_0 = 0$. When $i^\star > 1$, by the definition of $i^\star$, we have $Q'_-(\bar{\lambda}_{i^\star-1}) \geq 0$. By the optimality condition of concave functions, it is then clear that $\lambda^\star = \bar{\lambda}_{i^\star-1}$. On the other hand, if $Q'_+(\bar{\lambda}_{i^\star-1}) = -K + \Theta_{i^\star}(\bar{\lambda}_{i^\star-1}) > 0$, then since (i) $Q'_-(\bar{\lambda}_{i^\star}) < 0$ by the definition of $i^\star$ and (ii) $Q'(\lambda)$ is continuous in $(\bar{\lambda}_{i^\star-1}, \bar{\lambda}_{i^\star})$, there must exist some $\lambda^\star \in (\bar{\lambda}_{i^\star-1}, \bar{\lambda}_{i^\star})$ such that $Q'(\lambda^\star) = 0$ (*i.e.*, $\Theta_{i^\star}(\lambda^\star) = K$), concluding *Step 3*. □

Proof of Theorem 5. With the optimal dual variable $\lambda^\star$, we can determine the optimal order quantity leveraging (EC.21) and (EC.22) established in *Step 2* of the proof for Theorem 4. □

Proof of Theorem 6. By the interchangeability principle (see Lemma EC.3), we have

$$\max_{\psi\geq 0} \min_{G\in\mathcal{B}(\theta)} \min_{F\in\mathcal{P}} \{\mathbb{E}_F[\pi(\psi,\tilde{u})] + \alpha\cdot d(F,G)\} = \max_{\psi\geq 0} \min_{G\in\mathcal{B}(\theta)} \mathbb{E}_G\left[\min_{u\geq 0}\{\pi(\psi,u) + \alpha(u-\tilde{v})^2\}\right],$$

Using a standard duality argument (see, *e.g.*, Gao and Kleywegt 2023), the right-hand side problem becomes

$$\begin{aligned}&\max_{\psi\geq 0}\sup_{t\geq 0}\Big\{-t\theta + \mathbb{E}_H\Big[\min_{u\geq 0,\, v\geq 0}\{\pi(\psi,u) + \alpha(u-v)^2 + t(v-\tilde{w})^2\}\Big]\Big\}\\ =\ &\max_{\psi\geq 0}\sup_{t\geq 0}\Big\{-t\theta + \mathbb{E}_H\Big[\min_{u\geq 0}\{\pi(\psi,u) + \tfrac{\alpha t}{\alpha+t}(u-\tilde{w})^2\}\Big]\Big\},\end{aligned}$$

where the equality follows since $\min_{v\geq 0}\{\alpha(u-v)^2 + t(v-w)^2\} = \frac{\alpha t}{\alpha+t}(u-w)^2$ for any $u\geq 0$ and $w\geq 0$. Interchanging the "max" over ψ and t and applying the variable substitution $\gamma \leftarrow \frac{\alpha t}{\alpha+t}$, we arrive at

$$\max_{0\leq\gamma\leq\alpha}\Big\{-\tfrac{\alpha\gamma}{\alpha-\gamma}\theta + \max_{\psi\geq 0}\mathbb{E}_H\Big[\min_{u\geq 0}\{\pi(\psi,u) + \gamma(u-\tilde{w})^2\}\Big]\Big\}. \tag{EC.23}$$

Here, $\gamma = \alpha$ corresponds to $t = \infty$, and $\lim_{\gamma\to\alpha-} -\frac{\alpha\gamma}{\alpha-\gamma}\theta = 0$ if $\theta = 0$ and $\lim_{\gamma\to\alpha-} -\frac{\alpha\gamma}{\alpha-\gamma}\theta = \infty$ if $\theta > 0$. Hence, there must exists $\gamma^\star \in [0,\alpha]$ such that problem (EC.23) can be equivalently solved by

$$\begin{aligned}\max_{\psi\geq 0}\mathbb{E}_H\left[\min_{u\geq 0}\{\pi(\psi,u) + \gamma^\star\cdot(u-\tilde{w})^2\}\right] &= \max_{\psi\geq 0}\min_{F\in\mathcal{P},\,\Gamma\in\mathcal{W}(F,H)}\mathbb{E}_\Gamma[\pi(\psi,\tilde{u}) + \gamma^\star\cdot(\tilde{u}-\tilde{w})^2]\\ &= \max_{\psi\geq 0}\min_{F\in\mathcal{P}}\{\mathbb{E}_F[\pi(\psi,\tilde{u})] + \gamma^\star\cdot d(F,H)\},\end{aligned}$$

where the first equality follows from Lemma EC.3, and the second equality follows from the definition of $d(F,H)$. Hence, problem (13) is equivalent to $\max_{\psi\geq 0}\min_{F\in\mathcal{P}}\{\mathbb{E}_F[\pi(\psi,\tilde{u})] + \gamma^\star\cdot d(F,H)\}$ for some $\gamma^\star\in[0,\alpha]$.

In the remainder of the proof, we solve the optimal $\gamma^\star$ and $\psi^\star_{\gamma^\star}$ of problem (EC.23). For any fixed $w\geq 0$, define $\Psi(\gamma,q,w) = \min_{u\geq 0}\{\pi(q,u) + \gamma(u-w)^2\}$. Problem (EC.23) is then equivalent to

$$\max_{0\leq\gamma\leq\alpha}\Big\{-\tfrac{\alpha\gamma}{\alpha-\gamma}\theta + \max_{\psi\geq 0}\mathbb{E}_H\big[\Psi(\gamma,\psi,\tilde{w})\big]\Big\}. \tag{EC.24}$$

We first solve, by the first-order optimality condition, the inner maximization of problem (EC.24) given $\gamma\in[0,\alpha]$. Let $Z(\psi) = \mathbb{E}_H[\Psi(\gamma,\psi,\tilde{w})]$. Given $w\geq 0$, by the concavity of $\Psi(\gamma,\psi,w)$, $Z(\psi)$ is also concave with a decreasing derivative. In particular, if $\psi < \frac{p}{4\gamma}$, we have $Z'(\psi) = p\cdot\mathbb{P}_H\{\gamma\tilde{w}^2 \geq p\psi\} - c$; if $\psi \geq \frac{p}{4\gamma}$, we have $Z'(\psi) = p\cdot\mathbb{P}_H\{\tilde{w} - \frac{p}{4\gamma} \geq \psi\} - c$. Note that $Z'(0) = p - c > 0$ and $\lim_{\psi\to\infty} Z'(\psi) = -c < 0$. Let $\gamma_0 = \frac{p}{2q^\star_H}$. If $\gamma < \gamma_0$, then $Z'(\frac{p}{4\gamma}) = p(\kappa - H(\frac{p}{2\gamma})) < 0$. Hence, the maximum of $Z(\psi)$ is attained in $[0, \frac{p}{4\gamma}]$. Setting the derivative to 0 yields

$$\psi^\star_\gamma = (q^\star_H)^2\cdot\tfrac{\gamma}{p} < \tfrac{p}{4\gamma}. \tag{EC.25}$$

If $\gamma \geq \gamma_0$, then $Z'(\frac{p}{4\gamma}) = p(\kappa - H(\frac{p}{2\gamma})) \geq 0$, implying that the maximum of $Z(\psi)$ must be attained in $[\frac{p}{4\gamma}, \infty)$. Setting the derivative to 0 then yields

$$\psi^\star_\gamma = q^\star_H - \tfrac{p}{4\gamma} \geq \tfrac{p}{4\gamma}. \tag{EC.26}$$

We next solve the optimal $\gamma^\star$ of the outer maximization of problem (EC.24), whose objective is

$$Q(\gamma) = -\tfrac{\alpha\gamma}{\alpha-\gamma}\theta + \max_{\psi\geq 0}\mathbb{E}_H\big[\Psi(\gamma,\psi,\tilde{w})\big] = -\tfrac{\alpha\gamma}{\alpha-\gamma}\theta + \mathbb{E}_H\big[\Psi(\gamma,\psi^\star_\gamma,\tilde{w})\big].$$

Note that given $w\geq 0$, $\Psi(\gamma,\psi,w)$ is jointly concave in (γ,ψ). Since concavity is preserved under expectation and maximization, $Q(\gamma)$ is also concave in γ. When $\gamma < \gamma_0$, plugging the expressions of $\psi^\star_\gamma$ in (EC.25) and $\Psi(\alpha,\psi,w)$ in Lemma EC.3, we have

$$Q(\gamma) = -\tfrac{\alpha\gamma}{\alpha-\gamma}\theta + \mathbb{E}_H\big[\Psi(\gamma,\psi^\star_\gamma,\tilde{w})\big] = -\tfrac{\alpha\gamma}{\alpha-\gamma}\theta + \int_0^{q^\star_H} zw^2\mathrm{d}H(w),\quad Q'(\gamma) = -\tfrac{\alpha^2}{(\alpha-\gamma)^2}\theta + \int_0^{q^\star_H} w^2\mathrm{d}H(w).$$

When $\gamma \geq \gamma_0$, plugging the expression of $\psi^\star_\gamma$ in (EC.26), we have

$$Q(\gamma) = -\frac{\alpha\gamma}{\alpha-\gamma}\theta + \int_0^{\frac{p}{2\gamma}} \gamma w^2 \mathrm{d}H(w) + \int_{\frac{p}{2z}}^{q_H^\star} p(w - \tfrac{p}{4\gamma})\mathrm{d}H(w), \quad Q'(\gamma) = -\frac{\alpha^2\theta}{(\alpha-\gamma)^2} + \frac{p^2}{4\gamma^2}(\kappa - H(\tfrac{p}{2\gamma})) + \int_0^{\frac{p}{2\gamma}} w^2 \mathrm{d}H(w).$$

Note that $Q'(\gamma)$ is always decreasing in γ. When $\theta = 0$, it is straightforward to see that $Q'(\gamma) \geq 0$. Hence, $\gamma^\star = \alpha$ (which corresponds to $t^\star = \infty$ in the dual reformulation). In the following, we focus on $\theta > 0$. Based on the sign of $Q'(0)$, we divide the problem into two scenarios: $\theta \geq \beta$ and $\theta < \beta$, with $\beta = \int_0^{q_H^\star} u^2 \mathrm{d}H(u)$.

For the former scenario of $\theta \geq \beta$, $Q'(\gamma) \leq Q'(0) = -\theta + \beta \leq 0$ for any $\gamma \in [0, \alpha]$. Hence, the maximum of $Q(\gamma)$ is attained at $\gamma^\star = 0$. Equation (EC.25) then yields $\psi^\star_{\gamma^\star} = 0$. This corresponds to the case (i) in the statement. For the latter scenario of $\theta < \beta$, $Q'(0) = -\theta + \beta > 0$. To proceed, we further consider two situations based on the sign of $Q'(\gamma_0)$. If $1 - \sqrt{\theta/\beta} < \frac{\gamma_0}{\alpha}$, *i.e.*, $\sqrt{\theta/\beta} < \frac{\alpha}{\alpha-\gamma_0}$, then $Q'(\lambda_0) = \beta - \frac{\alpha^2}{(\alpha-\gamma_0)^2}\theta < 0$. Since $Q'(0) = -\theta + \beta > 0$, the maximum of $Q(\gamma)$ must be attained in $[0, \gamma_0]$. Setting the derivative of $Q(\gamma)$ to 0 yields $\gamma^\star = \alpha(1 - \sqrt{\theta/\beta}) < \gamma_0$. By equation (EC.25) we have $\psi^\star_{\gamma^\star} = (q_H^\star)^2(1 - \sqrt{\theta/\beta}) \cdot \frac{\alpha}{p}$. This corresponds to case (ii) in the statement. If $1 - \sqrt{\theta/\beta} \geq \frac{\gamma_0}{\alpha}$, then $Q'(\gamma_0) \geq 0$. Moreover, $\theta > 0$ gives $Q'_-(\alpha) = -\infty$, which further implies that there exists some $\gamma^\star \in [\gamma_0, \alpha)$ such that $-\frac{\alpha^2\theta}{(\alpha-\gamma^\star)^2} + \frac{p^2}{4\gamma^{\star 2}}\big(\kappa - H\big(\frac{p}{2\gamma^\star}\big)\big) + \int_0^{\frac{p}{2z^\star}} w^2 \mathrm{d}H(w) = 0$. Equation (EC.26) then yields $\psi^\star_{\gamma^\star} = q_H^\star - \frac{p}{4\gamma^\star}$. This corresponds to case (iii) in the statement. □

Proof of Theorem 7. If $\alpha = 0$, it is easy to see $q^\star_\alpha = 0$, and the result follows. We next focus on $\alpha > 0$. By the dual representation of total variation distance (see, *e.g.*, Müller 1997), we have $d_{\mathrm{TV}}(F, G) = \max_{\|\chi\|_\infty \leq 1}\{\mathbb{E}_F[\chi(v)] - \mathbb{E}_G[\chi(v)]\}$. Hence, for any $q \geq 0$ and $G \in \mathcal{A}$,

$$\begin{aligned}\min_{F \in \mathcal{P}}\{\mathbb{E}_F[\pi(q,\tilde{v})] + \alpha \cdot d_{\mathrm{TV}}(F,G)\} &= \max_{\|\chi\|_\infty \leq 1} \min_{F \in \mathcal{P}} \big\{\mathbb{E}_F[\pi(q,\tilde{v})] + \alpha\mathbb{E}_F[\chi(\tilde{v})] - \alpha\mathbb{E}_G[\chi(\tilde{v})]\big\} \\ &= \max_{\|\chi\|_\infty \leq 1} \Big\{\min_{v \geq 0}\{\pi(q,v)] + \alpha\chi(v)\} - \alpha\mathbb{E}_G[\chi(\tilde{v})]\Big\},\end{aligned}$$

where the first equality follows from the minimax theorem. Introducing an epigraphical variable τ, the above problem becomes

$$\begin{aligned}&\max_{\tau,\chi(\cdot)} \Big\{\tau - \alpha \cdot \mathbb{E}_G[\chi(\tilde{v})] \;\Big|\; \pi(q,v) + \alpha\chi(v) \geq \tau,\ |\chi(v)| \leq 1,\ \forall v \geq 0\Big\} \\ = &\max_{\tau} \big\{\tau - \alpha \cdot \min_{\chi(\cdot)} \mathbb{E}_G[\chi(\tilde{v})] \;\big|\; \chi(v) \geq \tfrac{\tau - \pi(q,v)}{\alpha},\ -1 \leq \chi(v) \leq 1,\ \forall v \geq 0\big\}.\end{aligned}$$

Given τ, we investigate the minimization of $\chi(\cdot)$. First, to ensure that the feasible set of $\chi(\cdot)$ is nonempty, we must have $\frac{\tau-\pi(q,v)}{\alpha} \leq 1$ for all $v \geq 0$, and hence $\tau \leq \alpha - cq$. Second, it suffices to focus on the pointwise minimum of $\chi(\cdot)$, and it is straightforward to see that the optimum is attained at $\chi^\star(v) = \max\{-1, \frac{\tau-\pi(q,v)}{\alpha}\}$ for any $v \geq 0$. Plugging in the expression of $\chi^\star$ into the objective, problem (15) then becomes

$$\max_{q \geq 0} \min_{G \in \mathcal{A}} \max_{\tau \leq \alpha - cq} \big\{\tau - \mathbb{E}_G[\max\{-\alpha, \tau - \pi(q,\tilde{v})\}].$$

We first look at the inner maximization over τ. For any $\tau \geq \alpha + cq$ (resp., $\tau \leq \alpha - (p-c)q$), $\min\{\alpha - \tau + cq, pv, pq\} = 0$ (resp., $= \min\{pv, pq\}$) and hence,, the maximum over τ must be attained in $[\alpha - (p-c)q, \alpha + cq]$. This, together with the prerequisite $\tau \geq cq - \alpha$, implies that it suffices to focus on $0 \leq \alpha - \tau + cq \leq \min\{pq, 2\alpha\}$.

We next look at the optimization over $q \geq 0$:

$$\begin{aligned}\max_{q \geq 0} \min_{G \in \mathcal{A}} \max_{0 \leq \alpha - \tau + cq \leq \min\{pq, 2\alpha\}} \{\mathbb{E}_G[\min\{\alpha - \tau + cq, p\tilde{v}\}] - cq\} &= \max_{q \geq 0} \min_{G \in \mathcal{A}} \max_{0 \leq \tau \leq \min\{pq, 2\alpha\}} \{\mathbb{E}_G[\min\{\tau, p\tilde{v}\}] - cq\} \\ &= \max_{q \geq 0} \min_{G \in \mathcal{A}} \{\mathbb{E}_G[\min\{pq, 2\alpha, p\tilde{v}\}] - cq\},\end{aligned}$$

where the first equality follows from the variable substitution $\tau \leftarrow \alpha - \tau + cq$ and the second equality follows from the fact that $\mathbb{E}_G[\min\{\tau, p\tilde{v}\}]$ is increasing in τ so its maximum is attained at $\min\{pq, 2\alpha\}$. When $q \geq \frac{2\alpha}{p}$, the objective function $\mathbb{E}_G[\min\{pq, 2\alpha, p\tilde{v}\}] - cq = \mathbb{E}_G[\min\{2\alpha, p\tilde{v}\}] - cq$ is decreasing in q given $G \in \mathcal{A}$. Hence, it is optimal to set q to $\frac{2\alpha}{p}$ whenever $q \geq \frac{2\alpha}{p}$. That is to say, the optimal order quantity of problem (15) must reside in $[0, \frac{2\alpha}{p}]$. For this interval, we have

$$\max_{0 \leq q \leq \frac{2\alpha}{p}} \min_{G \in \mathcal{A}} \{\mathbb{E}_G[\min\{pq, 2\alpha, p\tilde{v}\}] - cq\} = \max_{0 \leq q \leq \frac{2\alpha}{p}} \min_{G \in \mathcal{A}} \mathbb{E}_G[\pi(q,\tilde{v})].$$

Hence, it suffices to solve the right-hand side problem. We note that $\Phi(q) = \min_{G \in \mathcal{A}} \mathbb{E}_G[\pi(q,\tilde{v})]$ is concave in q. Hence, $q^\star_\alpha = \min\{\frac{2\alpha}{p}, q^\star_\infty\}$. Combining the expression of $q^\star_\infty$ in (1) then completes the proof. □

EC.3. Equivalence between $\mathcal{V}$ and $\mathcal{A}$ for Misspecification

We show that for Misspecification, it is equivalent to consider $\mathcal{V}=\{G\in\mathcal{P}\mid \mathbb{E}_G[\tilde v]=\mu,\ \mathbb{E}_G[\tilde v^2]=\mu^2+\sigma^2$ for some $\mu\in[\underline{\mu},\bar\mu]$ and $\sigma^2\in[\underline{\sigma}^2,\bar\sigma^2]\}$ with uncertain mean and variance and and $\mathcal{A}=\{G\in\mathcal{P}\mid \mathbb{E}_G[\tilde v]=\underline{\mu},\ \mathbb{E}_G[\tilde v^2]=\underline{\mu}^2+\bar\sigma^2\}$ with the lowest mean and the highest variance.

Lemma EC.5. *For any $\alpha\ge 0$, it holds that*

$$\max_{q\ge 0}\min_{F\in\mathcal{P}}\left\{\mathbb{E}_F[\pi(q,\tilde u)]+\alpha\cdot d(F,\mathcal{V})\right\}=\max_{q\ge 0}\min_{F\in\mathcal{P}}\left\{\mathbb{E}_F[\pi(q,\tilde u)]+\alpha\cdot d(F,\mathcal{A})\right\}.$$

Proof. By the definition of $\mathcal{V}$, for any $q\ge 0$, the left-hand side objective for a fixed q can be rewritten as

$$\min_{\mu\in[\underline{\mu},\bar\mu],\sigma^2\in[\underline{\sigma},\bar\sigma]}\ \min_{F\in\mathcal{P}}\left\{\mathbb{E}_F[\pi(q,\tilde u)]+\alpha\cdot d(F,\mathcal{A}(\mu,\sigma))\right\}=\min_{\mu\in[\underline{\mu},\bar\mu],\sigma^2\in[\underline{\sigma},\bar\sigma]} L(\mu,\sigma,q),$$

where $\mathcal{A}(\mu,\sigma)=\{G\in\mathcal{P}\mid \mathbb{E}_G[\tilde v]=\mu,\ \mathbb{E}_G[\tilde v^2]=\mu^2+\sigma^2\}$ and the expression of $L(\mu,\sigma;q)$ has been characterized in Proposition 1. Then it suffices to show that $L(\mu,\sigma,q)$ is always increasing in μ and decreasing in σ. Note that the function $L(\mu,\sigma,q)$ consists of two pieces: $L_1(\mu,\sigma,q)=\frac{p}{2}(q+\mu-\frac{p}{4\alpha}-\sqrt{(q-\mu+\frac{p}{4\alpha})^2+\sigma^2})-cq$ and $L_2(\mu,\sigma,q)=\frac{\alpha}{2}(\frac{pq}{\alpha}+\mu^2+\sigma^2-\sqrt{(\frac{pq}{\alpha}+\mu^2+\sigma^2)^2-4\mu^2\frac{pq}{\alpha}})-cq$. On the one hand, it is straightforward to see that $L_1(\mu,\sigma,q)$ is decreasing in σ, and we also have $\frac{\partial L_1(\mu,\sigma,q)}{\partial\mu}=\frac{p}{2}\big(1+\frac{q-\mu+\frac{p}{4\alpha}}{\sqrt{(q-\mu+\frac{p}{4\alpha})^2+\sigma^2}}\big)\ge 0$. On the other hand, we note that $L_2(\mu,\sigma,p)=\frac{\alpha}{2}\frac{4\mu^2\frac{pq}{\alpha}}{\frac{pq}{\alpha}+\mu^2+\sigma^2+\sqrt{(\frac{pq}{\alpha}+\mu^2+\sigma^2)^2-4\mu^2\frac{pq}{\alpha}}}$ is decreasing in σ. Additionally, $\frac{\partial L_2(\mu,\sigma,q)}{\partial\mu}=\alpha\mu\big(1-\frac{\frac{pq}{\alpha}+\mu^2+\sigma^2-2\frac{pq}{\alpha}}{\sqrt{(\frac{pq}{\alpha}+\mu^2+\sigma^2)^2-4\mu^2\frac{pq}{\alpha}}}\big)\ge 0$. Finally, one can easily verify that $L(\mu,\sigma,q)$ is continuous in μ and σ, and hence, $L(\mu,\sigma,q)$ is increasing in μ and decreasing in σ, concluding the proof. □

EC.4. Transformed Worst-Case Distribution

In this section, we derive the worst-case transformed distribution $T_{\varphi_\alpha}[G^\star_\alpha]$ given an order quantity $q\ge 0$. For ease of notation, let $w(q)=\frac{pq}{\alpha}+\mu^2+\sigma^2$, $u(q)=q+\frac{p}{4\alpha}$, and define

$$\begin{cases} v_1=\frac{1}{2\mu}\big(w(q)-\sqrt{w(q)^2-4\mu^2\frac{pq}{\alpha}}\big),\ v_2=\frac{1}{2\mu}\big(w(q)+\sqrt{w(q)^2-4\mu^2\frac{pq}{\alpha}}\big)\\ v_3=u(q)-\sqrt{(u(q)-\mu)^2+\sigma^2},\ v_4=u(q)+\sqrt{(u(q)-\mu)^2+\sigma^2}.\end{cases}$$

Proposition EC.1 (Worst-Case Transformed Distribution). *Given $q\ge 0$ and $\alpha\ge 0$, the worst-case transformed distribution $T_{\varphi_\alpha}[G^\star_\alpha]$ of Transform is given by*

$$T_{\varphi_\alpha}[G^\star_\alpha]=\begin{cases}\left(\frac12-\frac{\mu^2-\sigma^2-\frac{pq}{\alpha}}{2\sqrt{w(q)^2-4\mu^2\frac{pq}{\alpha}}}\right)\cdot\delta_{\frac{\alpha}{p}v_1^2}+\left(\frac12+\frac{\mu^2-\sigma^2-\frac{pq}{\alpha}}{2\sqrt{w(q)^2-4\mu^2\frac{pq}{\alpha}}}\right)\cdot\delta_{\frac{\alpha}{p}v_2^2} & 0\le q\le\frac{p}{4\alpha}\\ \left(\frac12-\frac{\mu^2-\sigma^2-\frac{pq}{\alpha}}{2\sqrt{w(q)^2-4\mu^2\frac{pq}{\alpha}}}\right)\cdot\delta_{\frac{\alpha}{p}v_1^2}+\left(\frac12+\frac{\mu^2-\sigma^2-\frac{pq}{\alpha}}{2\sqrt{w(q)^2-4\mu^2\frac{pq}{\alpha}}}\right)\cdot\delta_{v_2-\frac{p}{4\alpha}} & q\ge\frac{p}{4\alpha},(2\mu-\frac{p}{\alpha})q<\mu^2+\sigma^2-\frac{p\mu}{2\alpha}\\ \frac12\left(1+\frac{u(q)-\mu}{\sqrt{(u(q)-\mu)^2+\sigma^2}}\right)\cdot\delta_{v_3-\frac{p}{4\alpha}}+\frac12\left(1-\frac{u(q)-\mu}{\sqrt{(u(q)-\mu)^2+\sigma^2}}\right)\cdot\delta_{v_4-\frac{p}{4\alpha}} & q\ge\frac{p}{4\alpha},(2\mu-\frac{p}{\alpha})q<\mu^2+\sigma^2-\frac{p\mu}{2\alpha}.\end{cases}$$

Proof. Given $q\ge 0$, we have derived the expression of the worst-case distribution $G^\star_\alpha$ as in (EC.5) and (EC.7). Plugging the expression of φ_α, we then obtain the desired results. □

In Figure EC.2, we provide a visualization of the transformed worst-case distribution $T_{\varphi_\alpha}[G^\star_\alpha]$, pertaining to the worst-case distribution $G^\star_\infty$ for Ambiguity and the nominal distribution as truncated normal. Define $v_1^\star=\mu-\sigma\sqrt{\frac{c}{p-c}}$ and $v_2^\star=\mu+\sigma\sqrt{\frac{p-c}{c}}$. Plugging in the expression of $q^\star_\alpha$ in Theorem 2, the worst-case transformed distribution $T_{\varphi_\alpha}[G^\star_\alpha]$ can be characterized as follows.

(*i*) If $\frac{p-c}{p}<\frac{\sigma^2}{\mu^2+\sigma^2}$, then $T_{\varphi_\alpha}[G^\star_\alpha]=\big(\frac{\sigma^2}{\mu^2+\sigma^2}\big)\cdot\delta_0+\big(\frac{\mu^2}{\mu^2+\sigma^2}\big)\cdot\delta_{\frac{\mu^2+\sigma^2}{\mu}}$.

(*ii*) If $\frac{p-c}{p}\ge\frac{\sigma^2}{\mu^2+\sigma^2}$, then

$$T_{\varphi_\alpha}[G^\star_\alpha]=\begin{cases}\big(\frac{p-c}{p}\big)\cdot\delta_{\frac{\alpha}{p}v_1^{\star 2}}+\big(\frac{c}{p}\big)\cdot\delta_{\frac{\alpha}{p}v_2^{\star 2}} & \alpha<p/(2\sqrt{v_1^\star v_2^\star})\\ \big(\frac{p-c}{p}\big)\cdot\delta_{\frac{\alpha}{p}v_1^{\star 2}}+\big(\frac{c}{p}\big)\cdot\delta_{v_2^\star-\frac{p}{4\alpha}} & p/(2\sqrt{v_1^\star v_2^\star})\le\alpha<p/(2v_1^\star)\\ \big(\frac{p-c}{p}\big)\cdot\delta_{v_1^\star-\frac{p}{4\alpha}}+\big(\frac{c}{p}\big)\cdot\delta_{v_2^\star-\frac{p}{4\alpha}} & \alpha\ge p/(2v_1^\star).\end{cases}$$

Setting $\alpha\to\infty$ then yields the expression of the worst-case distribution of Ambiguity: $G^\star_\infty=(\frac{\sigma^2}{\mu^2+\sigma^2})\cdot\delta_0+(\frac{\mu^2}{\mu^2+\sigma^2})\cdot\delta_{\frac{\mu^2+\sigma^2}{\mu}}$ if $\frac{p-c}{p}<\frac{\sigma^2}{\mu^2+\sigma^2}$ and $G^\star_\infty=(\frac{p-c}{p})\cdot\delta_{v_1^\star}+(\frac{c}{p})\cdot\delta_{v_2^\star}$ otherwise. It is evident that $G^\star_\infty$ and $T_{\varphi_\alpha}[G^\star_\alpha]$ are both two-point distributions with the same probability mass, and the minimal (resp., maximal) support point of $T_{\varphi_\alpha}[G^\star_\alpha]$ is always smaller than that of $G^\star_\infty$.

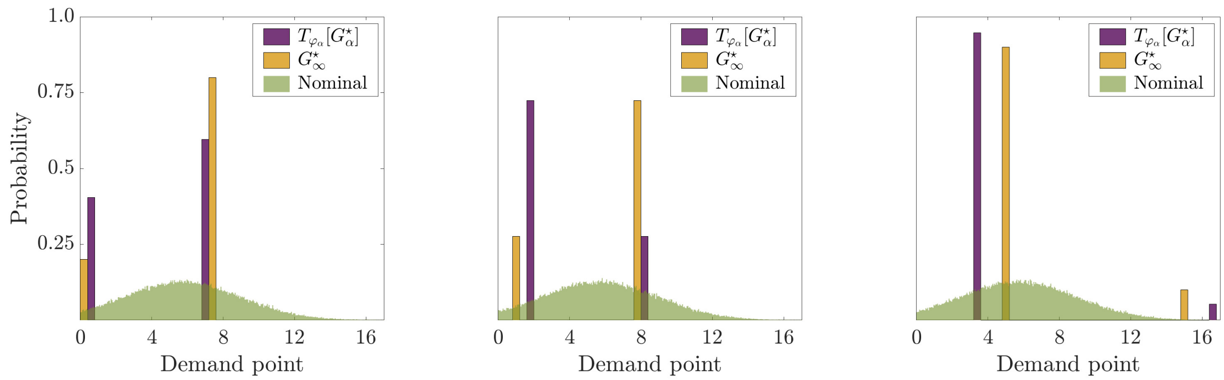


Figure EC.2 Three scenarios depending on the value of q: $0 \le q \le \frac{p}{4\alpha}$ (left), $q \ge \frac{p}{4\alpha}$ and $(2\mu - \frac{p}{\alpha})q < \mu^2 + \sigma^2 - \frac{p\mu}{2\alpha}$ (middle), and $q \ge \frac{p}{4\alpha}$ and $(2\mu - \frac{p}{\alpha})q \ge \mu^2 + \sigma^2 - \frac{p\mu}{2\alpha}$ (right).

EC.5. Generality of Theorem 5

We show how Theorem 5 generalizes Theorem 2 for a single product to multiple products. To see this, let $K = \mu^2 + \sigma^2$ and note that for the single-product problem, $\lambda^\star$ is the optimal dual variable $r^\star_\alpha$ in the dual reformulation. Plugging the expression of $q^\star_\alpha$ into equations (EC.6) and (EC.8), it is immediate to see that

$$r^\star_\alpha = \begin{cases} \sqrt{\kappa(1-\kappa)}/(2p\sigma) & \kappa \ge \frac{\sigma^2}{\mu^2+\sigma^2},\ \alpha \ge p/(2(\mu - \sigma\sqrt{(1-\kappa)/\kappa})) \\ \alpha\sqrt{\kappa(1-\kappa)}\big(\mu - \sigma\sqrt{(1-\kappa)/\kappa}\big)/\sigma & \kappa \ge \frac{\sigma^2}{\mu^2+\sigma^2},\ \alpha < \frac{p}{2(\mu-\sigma\sqrt{(1-\kappa)/\kappa})} \\ 0 & \kappa < \frac{\sigma^2}{\mu^2+\sigma^2}. \end{cases}$$

If $\kappa \ge \frac{\sigma^2}{\mu^2+\sigma^2}$ and $\alpha \ge \frac{p}{2(\mu-\sigma\sqrt{(1-\kappa)/\kappa})}$, it holds that $\lambda^\star = \frac{\sqrt{\kappa(1-\kappa)}}{2p\sigma}$ and hence, $\frac{p}{2(\mu - c/(2\lambda^\star))^+} = \frac{p}{2(\mu-\sigma\sqrt{(1-\kappa)/\kappa})}$. By Theorem 5, we have $q^\star_\alpha = \mu + \frac{p-2c}{4\lambda^\star} - \frac{p}{4\alpha} = \mu + \sigma\frac{2\kappa-1}{2\sqrt{\kappa(1-\kappa)}} - \frac{p}{4\alpha}$. If $\kappa \ge \frac{\sigma^2}{\mu^2+\sigma^2}$ and $\alpha < \frac{p}{2(\mu-\sigma\sqrt{(1-\kappa)/\kappa})}$, then $\lambda^\star = \frac{\alpha\sqrt{\kappa(1-\kappa)}}{\sigma}(\mu - \sigma\sqrt{(1-\kappa)/\kappa})$. In this case, we can verify that $\alpha < \frac{p}{2(\mu-c/(2\lambda^\star))^+}$, thus by Theorem 5, $q^\star_\alpha = \frac{\lambda^\star(\lambda^\star+\alpha)p\mu^2}{\alpha(\lambda^\star p/\alpha + c)^2} = (\mu^2 - \sigma^2 + 2\mu\sigma f(1-\kappa)) \cdot \frac{\alpha}{p}$. If $\kappa < \frac{\sigma^2}{\mu^2+\sigma^2}$, we always have $\lambda^\star = 0$, leading to $q^\star_\alpha = 0$. Consolidating these three scenarios then recovers the optimal solution (6) in Theorem 2.

EC.6. Multiple Products with Complete Covariance Information

In line with the setting of Section 6.1, we study misspecification-averse multi-product newsvendor with mean and complete covariance information. Consider M products with random demands $\tilde{\boldsymbol{u}} = (\tilde{u}_1, \ldots, \tilde{u}_i) \sim F$ following a multi-dimensional distribution F. The misspecification-averse newsvendor then solves

$$\max_{\boldsymbol{q} \ge \boldsymbol{0}} \min_{F \in \mathcal{P}_M} \{\mathbb{E}_F[\omega(\boldsymbol{q}, \tilde{\boldsymbol{u}})] + \alpha \cdot d(F, \mathcal{A})\}, \tag{EC.27}$$

where the function $\omega(\boldsymbol{q}, \boldsymbol{u}) = \sum_{i=1}^M \pi_i(q_i, u_i)$, the optimal-transport cost $d(\cdot,\cdot)$ is defined in (2) with L_2-norm, and the ambiguity set is specified by mean and complete covariance information, *i.e.*,

$$\mathcal{A} = \{G \in \mathcal{P}_M \mid \mathbb{E}_G[\tilde{\boldsymbol{v}}] = \boldsymbol{\mu},\ \mathbb{E}_G[(\tilde{\boldsymbol{v}} - \boldsymbol{\mu})(\tilde{\boldsymbol{v}} - \boldsymbol{\mu})^\top] = \boldsymbol{\Sigma}\}. \tag{EC.28}$$

Proposition EC.2. *Let $\mathcal{S}^M$ be the space of symmetric matrix with dimension M. For the ambiguity set (EC.28) with mean and covariance information, the misspecification-averse multi-product newsvendor problem (EC.27) is equivalent to*

$$\begin{aligned} \max\ & t + \boldsymbol{\lambda}^\top \boldsymbol{\mu} + \langle \boldsymbol{Q}, \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\top \rangle \\ \text{s.t.}\ \ & t + \boldsymbol{\lambda}^\top \boldsymbol{v} + \langle \boldsymbol{Q}, \boldsymbol{v}\boldsymbol{v}^\top \rangle \le \sum_{i \in [M]} \Psi_i(\alpha, q_i, v_i)\ \forall \boldsymbol{v} \ge 0 \\ & \boldsymbol{q} \ge 0,\ \boldsymbol{Q} \in \mathcal{S}^M,\ \boldsymbol{\lambda} \in \mathbb{R}^M,\ t \in \mathbb{R}, \end{aligned} \tag{EC.29}$$

where for each $i \in [M]$, $\Psi_i(\alpha, q_i, v_i) = \min\{\alpha v_i^2, p q_i\} - c_i q_i$ if $0 \le q_i \le \frac{p_i}{4\alpha}$, and if $q_i > \frac{p_i}{4\alpha}$,

$$\Psi_i(\alpha, q_i, v_i) = \begin{cases} \alpha v_i^2 - c_i q_i & 0 \le v_i \le \frac{p_i}{2\alpha} \\ p \cdot \min\{v_i - \frac{p_i}{4\alpha}, q_i\} - c_i q_i & v_i > \frac{p_i}{2\alpha}. \end{cases}$$

Proof. For $\boldsymbol{q} \geq 0$, by the interchangeability principle in Lemma EC.3, we have

$$\max_{\boldsymbol{q}\geq 0} \min_{G\in\mathcal{C}} \min_{F\in\mathcal{P}_M} \left\{\mathbb{E}_F[\omega(\boldsymbol{q},\tilde{\boldsymbol{u}})] + \alpha\cdot d(F,G)\right\} = \max_{\boldsymbol{q}\geq 0} \min_{G\in\mathcal{C}} \mathbb{E}_G\left[\min_{\boldsymbol{u}\geq 0}\{\omega(\boldsymbol{q},\tilde{\boldsymbol{u}}) + \alpha\|\boldsymbol{u}-\tilde{\boldsymbol{v}}\|_2^2\}\right],$$

which, by noting that the inner objective is separable, further reduces to

$$\max_{\boldsymbol{q}\geq 0} \min_{G\in\mathcal{C}} \mathbb{E}_G\left[\sum_{i\in[M]} \min_{u_i\geq 0}\{\pi(q_i,u_i) + \alpha(u_i-\tilde{v}_i)^2\}\right] = \max_{\boldsymbol{q}\geq 0} \min_{G\in\mathcal{C}} \mathbb{E}_G\left[\sum_{i\in[M]} \Psi_i(\alpha,q_i,\tilde{v}_i)\right]. \tag{EC.30}$$

Here, the expression of $\Psi_i(\alpha,q_i,\tilde{v}_i)$ is given in Lemma EC.4. Applying the standard duality argument for the moment problem, we then obtain the desired result. $\square$

Recall from Natarajan et al. (2018) that the ambiguity-averse multi-product newsvendor problem

$$\max_{\boldsymbol{q}\geq \boldsymbol{0}} \min_{G\in\mathcal{A}} \mathbb{E}_G[\omega(\boldsymbol{q},\tilde{\boldsymbol{v}})]$$

admits an equivalent dual reformulation

$$\begin{aligned} \max\ & t + \boldsymbol{\lambda}^\top\boldsymbol{\mu} + \langle \boldsymbol{Q}, \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\top\rangle \\ \text{s.t.}\ & t + \boldsymbol{\lambda}^\top\boldsymbol{v} + \langle \boldsymbol{Q}, \boldsymbol{v}\boldsymbol{v}^\top\rangle \leq \sum_{i\in[M]} \pi_i(q_i,v_i) \quad \forall \boldsymbol{v}\geq 0 \\ & \boldsymbol{q}\geq 0,\ \boldsymbol{Q}\in\mathcal{S}^M,\ \boldsymbol{\lambda}\in\mathbb{R}^M,\ t\in\mathbb{R}, \end{aligned} \tag{EC.31}$$

which involves 2^M quadratic constraints by a complete expansion of the function $\sum_{i\in[M]}\pi_i(q_i,v_i)$, and is already known to be intractable due to the full covariance structure (Hanasusanto et al. 2015). Clearly, the function $\Psi_i(\alpha,q_i,v_i)$ in (EC.29) is more complicated than the newsvendor profit function $\pi_i(q_i,v_i)$ as the former depends on the value of q_i (see Figure EC.1 for a visualization). Therefore, the misspecification-averse problem (EC.29), compared to (EC.31), is less tractable and more challenging to solve. Specifically, to solve this problem, we first need to partition the decision space into 2^M subregions (according to $q_i \leq \frac{p_i}{4\alpha}$ or not), where in each subregion, the function form of $\Psi_i(\alpha,q_i,v_i)$ is fixed. We then solve these 2^M subproblems, where for each subproblem, there are at least 2^M quadratic constraints in the dual reformulation by expanding $\sum_{i\in[M]}\Psi_i(\alpha,q_i,v_i)$. Therefore, solving a multi-product misspecification-averse problem is much harder than its ambiguity-averse counterpart.

To resolve this issue, we can identify a lower bound of (EC.30), by noting that for each $i\in[M]$,

$$\Psi_i(\alpha,q_i,v_i) \geq p_i\cdot\min\left\{q_i, v_i - \frac{p_i}{4\alpha}\right\} - c_iq_i = L_i(\alpha,q_i,v_i),\ \forall q_i\geq 0,\ v_i\geq 0.$$

Replacing $\Psi_i(\alpha,q_i,v_i)$ with $L_i(\alpha,q_i,v_i)$ in problem (EC.30), we then obtain a lower bound. The function $L_i(\alpha,q_i,v_i)$ inherits a similar structure to that of $\pi(q_i,v_i)$. Hence, the resulting problem can also be approximately solved via the quadratic decision rules or semi-definite programming relaxations as discussed in Hanasusanto et al. (2015), Natarajan and Teo (2017) and Natarajan et al. (2018).

EC.7. Calibrating α with Limited Knowledge of Distribution Shift

Assume the access to training samples $\hat{v}_1,\ldots,\hat{v}_N$ drawn from the data-generating distribution D and a very small number of testing samples $\hat{u}_1,\ldots,\hat{u}_M$ drawn from the out-of-sample distribution F, with $M \ll N$ (in our experiments, $N=30$ and $M=5$). The corresponding empirical distributions are $\hat{D} = \frac{1}{N}\sum_{n=1}^N \delta_{\hat{v}_n}$ and $\hat{F} = \frac{1}{M}\sum_{m=1}^M \delta_{\hat{u}_m}$. This allows us to compute the empirical distribution shift $d(\hat{F},\hat{D})$ used to estimate the underlying $d(F,D)$.

We next describe in detail the two strategies for calibrating α based on the testing samples.

- **Formula-based approach**. This approach explores formula (8) derived in Theorem 3, *i.e.*, $\alpha_N = \frac{1}{2}\sqrt{\frac{p(p-c)}{\epsilon_N + d(F,D)}}$, to determine α. To implement the formula in practice, it remains to estimate the unknown quantities $d(F,D)$ and ϵ_N, which correspond to the following two steps accordingly.

  Step 1. We first employ the term $\beta\cdot d(\hat{F},\hat{D})$ as an estimate for the distance term $d(F,D)$, where $d(\hat{F},\hat{D})$ is the empirical distribution shift as computed before and β is randomly drawn from $\mathbb{U}[0.5,1]$ as a randomized discount factor.

  Step 2. The quantity ϵ_N involves unknown problem-specific constants in its explicit form given in Theorem 3 and is thus not directly computable. Therefore, to select α_N, we treat ϵ_N as a tuning parameter and perform a 5-fold cross-validation for ϵ_N over an interval $[\underline{\epsilon},\bar{\epsilon}]$ using the training distribution $\hat{D}$. This is actually equivalent to cross-validation of α_N over the interval $\left[\frac{1}{2}\sqrt{\frac{p(p-c)}{\bar{\epsilon}+d(F,D)}}, \frac{1}{2}\sqrt{\frac{p(p-c)}{\underline{\epsilon}+d(F,D)}}\right]$.

- **Stress-testing approach**. This approach can be viewed as a *controlled stress test* (Kupiec 2002) that proceeds with two steps: first, constructing a stress-testing distribution; and second, validating based on the constructed stress-testing distribution.

  Step 1. To mimic the testing environment with distribution shift, we generate a synthetic stress-testing distribution F_{stress} supported on N points that satisfies $d(F_{\text{stress}}, \hat{D}) = \beta \cdot d(\hat{F}, \hat{D})$, where $\beta \in [0.5, 1]$ is a discount factor that downscales the empirical distribution shift to account for potential overestimation arising from the imbalance in sample sizes between $\hat{D}$ and $\hat{F}$. Here, we randomly choose the discount factor β from a uniform distribution $\mathbb{U}[0.5, 1]$. To achieve this, one admissible construction is $F_{\text{stress}} = \frac{1}{N}\sum_{n=1}^{N} \delta_{\hat{w}_n}$ with $\hat{w}_n = (1-\rho)\hat{v}_n + \rho\hat{v}_*$, where $\hat{v}_* = \min_{n\in[N]} \hat{v}_n$ and $\rho = \sqrt{\frac{N\beta d(\hat{F},\hat{D})}{\sum_{n\in[N]}(\hat{v}_n - \hat{v}_*)^2}}$. Here, ρ is the stress multiplier that determines how much mass is shifted downward so that F_{stress} exactly matches the empirical distance $\beta d(\hat{F}, \hat{D})$ from $\hat{D}$. By construction, one can easily verify that the constructed F_{stress} satisfies $d(F_{\text{stress}}, \hat{D}) = \frac{1}{N}\sum_{n\in[N]}(\hat{v}_n - \hat{w}_n)^2 = \beta d(\hat{F}, \hat{D})$, and the resultant samples of F_{stress} are shifted downward relative to those of the training distribution $\hat{D}$, which is then used for validating α in a distribution shifted environment.

  Step 2. Furthermore, based on the empirical mean $\hat{\mu}$ and standard deviation $\hat{\sigma}$ of $\hat{D}$, the hyperparameter α is then cross-validated over an interval $[\underline{\alpha}, \bar{\alpha}]$ using the generated testing distribution F_{stress}.

References

Besbes, Omar, Will Ma, Omar Mouchtaki. 2023. From contextual data to newsvendor decisions: On the actual performance of data-driven algorithms. *Available at* https://arxiv.org/abs/2302.08424.

Dowson, D.C., B.V. Landau. 1982. The Fréchet distance between multivariate normal distributions. *Journal of Multivariate Analysis* **12**(3) 450–455.

Gao, Rui, Anton Kleywegt. 2023. Distributionally robust stochastic optimization with Wasserstein distance. *Mathematics of Operations Research* **48**(2) 603–655.

Gelbrich, Matthias. 1990. On a formula for the L^2 Wasserstein metric between measures on Euclidean and Hilbert spaces. *Mathematische Nachrichten* **147**(1) 185–203.

Hanasusanto, Grani A., Daniel Kuhn, Stein W. Wallace, Steve Zymler. 2015. Distributionally robust multi-item newsvendor problems with multimodal demand distributions. *Mathematical Programming* **152**(1) 1–32.

Kupiec, Paul. 2002. Stress testing in a value at risk framework. *Risk management: value at risk and beyond*, 76–99. Cambridge University Press.

Müller, Alfred. 1997. Integral probability metrics and their generating classes of functions. *Advances in Applied Probability* **29**(2) 429–443.

Natarajan, Karthik, Chung-Piaw Teo. 2017. On reduced semidefinite programs for second order moment bounds with applications. *Mathematical Programming* **161** 487–518.

Natarajan, Karthik, Melvyn Sim, Joline Uichanco. 2018. Asymmetry and ambiguity in newsvendor models. *Management Science* **64**(7) 3146–3167.

Rigollet, Philippe, Hütter Jan-Christian. 2023. High-dimensional statistics. *Available at* https://arxiv.org/abs/2310.19244.

Scarf, Herbert. 1958. A min-max solution of an inventory problem. *Studies in the mathematical theory of inventory and production.*

Villani, Cédric. 2009. *Optimal transport: Old and new.* Springer.

Wainwright, Martin. 2019. High-dimensional statistics: A non-asymptotic viewpoint. *Cambridge University Press.*

Zhang, Luhao, Jincheng Yang, Rui Gao. 2025. A short and general duality proof for Wasserstein distributionally robust optimization. *Operations Research* **73**(4) 2146–2155.